MISSING RESPONSES AND IRT ABILITY ESTIMATION: OMITS, CHOICE, TIME LIMITS, AND ADAPTIVE TESTING

Robert J. Mislevy
Pao-Kuei Wu

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 1996	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE
Missing responses and IRT Ability Estimation:
Omits, choice, time limits, and adaptive testing.

5. FUNDING NUMBERS
G.   N00014-91-J-4101
PE. 61153N
PR. RR 04204
TA. RR 04204-01
WU. R & T 4421537-01

6. AUTHOR(S)
Robert J. Mislevy
Pao-kuei Wu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road
Princeton, NJ   08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-96-30-ONR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA   22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*

The basic equations of item response theory (IRT) provide a foundation for inferring examinees' abilities from responses to different test items. In practice, examinees do not generally provide a response to all items--for reasons that may or may not have been intended by the test administrator, and that may or may not be related to their ability. The mechanisms that produce missingness must be taken into account if correct inferences are to be drawn. Using concepts introduced by Rubin (1976), we discuss the implications for Bayesian and direct likelihood ability parameter estimation that are entailed by alternate test forms, targeted testing, adaptive testing, time limits, omitted responses, and examinee choice of tasks. Attention is focused on whether, in each case, the mechanism for missingness is "ignorable," and, in those cases in which it is not, how it can be modeled.

14. SUBJECT TERMS
Adaptive testing; choice; customized tests; item response theory; missing data; omitted responses; targeted testing.

15. NUMBER OF PAGES
50

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

Missing Responses and IRT Ability Estimation:

Omits, Choice, Time Limits, and Adaptive Testing

Robert J. Mislevy & Pao-Kuei Wu

Educational Testing Service

June, 1996

Missing Responses and IRT Ability Estimation:
Omits, Choice, Time Limits, and Adaptive Testing

Abstract

The basic equations of item response theory (IRT) provide a foundation for
inferring examinees' abilities from responses to different test items. In practice, examinees
do not generally provide a response to all items—for reasons that may or may not have
been intended by the test administrator, and that may or may not be related to their ability.
The mechanisms that produce missingness must be taken into account if correct inferences
are to be drawn. Using concepts introduced by Rubin (1976), we discuss the implications
for Bayesian and direct likelihood ability parameter estimation that are entailed by alternate
test forms, targeted testing, adaptive testing, time limits, omitted responses, and examinee
choice of tasks. Attention is focused on whether, in each case, the mechanism for
missingness is "ignorable," and, in those cases in which it is not, how it can be modeled.

1.0 Introduction

Item response theory (IRT) models the probability of an examinee's responses to test items as conditionally independent, given an unobservable ability parameter θ (Lord, 1980). The oft-cited capacity of IRT for measuring different examinees with different test items implies inference in the presence of missing data, since an examinee may not have provided a response to every item in the item domain of interest. The following types of missingness are in fact routinely encountered in applications of IRT:

- *Alternate test forms.* Two or more tests with similar content but different items are often employed to minimize carry-over effects, reduce fatigue and practice effects, or avoid cheating. An examinee is administered one form selected at random.

- *Targeted testing.* Tests pitched at different levels of difficulty make measurement more efficient when background information related to ability, such as grade or courses taken, can be used to determine which test to administer to each examinee.

- *Adaptive testing.* Testing can also be made more efficient if each item presented to an examinee is selected in light of responses thus far.

- *Not-reached items.* Under typical testing conditions, some examinees will not reach the last items on a test because of the time limit.

- *Omitted items.* Even when an item has been presented and an examinee has time to consider it, the examinee will sometimes choose not to respond.

- *Examinee choice.* Examinees may be allowed to examine a number of items, and choose which to answer, subject to specified constraints (e.g., "Answer any two of the following four questions").

When incomplete data are encountered, the IRT model that determines responses is embedded in a more encompassing model that determines which responses will be observed and which will be missing. This paper discusses the implications that missing responses hold for direct likelihood and Bayesian inferences about examinee ability parameters, assuming item parameters are known. When can the process that causes missingness be ignored? When it cannot be ignored, how can it be modeled? How can conventional IRT methods for missing responses be evaluated in this framework? Section 2 extends IRT notation to handle missingness, using concepts and notation from Little and Rubin (1987) and Rubin (1976). Next, Rubin's (1976) conditions for when the missingness process can be ignored are reviewed. Sections 3-8 address the six types of missingness listed above. Section 9 is a non-technical summary of the main results.

2.0   Background

Section 2.1 gives notation for IRT. Section 2.2 extends IRT to missing data situations using concepts and notation from Rubin (1976) and Little and Rubin (1987), and Section 2.3 lists Rubin's results on ignorability.

2.1   Notation for IRT

DEFINITION. An IRT model with examinee parameter θ is said to satisfy **local independence (LI)** in a domain of n items if

$$\text{Prob}(U_1 = u_1,\ldots,U_n = u_\text{n}|\theta,\beta_1,\ldots,\beta_n,y) = \prod_{j=1}^{n} \text{Prob}(U_j = u_j|\theta,\beta_j),$$

or, written more compactly,

$$\text{Prob}(U = u|\theta,\beta,y) = f_\theta(u) = \prod_{j=1}^{n} f_\theta(u_j), \tag{2.1}$$

where U_j is the response variable for Item j, u_j represents a value thereof, and $u = (u_1,\ldots,u_n)$; $f_\theta(\cdot)$ is the response function, interpreted as applying to individual items or sets of items in accordance with its arguments; β_j is a possibly vector-valued parameter characterizing the dependency of response probabilities to Item j on θ, and $\beta = (\beta_1,\ldots,\beta_n)$; and y denotes covariate information about examinees, such as age or courses taken.

It will be seen below that results for alternate test forms, targeted testing, adaptive testing, and not-reached items can be obtained without further specification of models in addition to the IRT model. Results for omitted items and choice items, however, require speculations about, and modeling of, the examinees' perspective on the missingness process. Sections dealing with these cases focus on a class of IRT models that is common in educational testing, namely those satisfying local independence, unidimensionality, and monotonicity (Holland & Rosenbaum, 1986). We adapt an acronym from Zwick (1990):

DEFINITION. An IRT model is said to be **SMURFLI**[2] if it satisfies the following conditions:

- Strict Monotonicity; i.e., $\theta' > \theta'' \Rightarrow \text{Prob}(U_j = 1|\theta') > \text{Prob}(U_j = 1|\theta'')$.

- Unidimensional Response Functions; i.e., the domain of θ is $\Re^1$.

- Local Independence.

- 2 possible responses, correct or incorrect, can be observed for each item; specifically, $u_j = 1$ indicates a correct answer and 0 an incorrect answer.

Under the Rasch model for dichotomous items, for example, θ and β_j are real numbers, and $f_\theta(u_j) = \exp[u_j(\theta - \beta_j)]/[1 + \exp(\theta - \beta_j)]$ (Figure 1).

[Figure 1 about here]

If there is no possibility of missing responses, (2.1) is interpreted as a likelihood function, say $L(\theta|\tilde{u})$, once a particular value $\tilde{u}$ of U has been observed. *Direct likelihood inferences* are based solely on relative values of L at different values of θ. One might say, for example, that the probability of $\tilde{u}$ at θ' is twice that at θ'', or that it attains its maximum at $\hat{\theta}$, the maximum likelihood estimate (MLE). Under direct likelihood inference, the MLE is interpreted only as a feature of the likelihood induced by the data that were actually observed, not as a realized value of an estimator with a reference distribution concerning repeated samples of U with a fixed "true" θ. This presentation does not focus on *sampling distribution inferences*, although some remarks will be made in passing. *Bayesian inferences* are based on the posterior distribution for θ given $\tilde{u}$, or

$$p(\theta|\tilde{u}) = K(\tilde{u})\, L(\theta|\tilde{u}) p(\theta), \qquad (2.2)$$

where $K(\tilde{u})$ is a normalizing constant and $p(\theta)$ is the prior distribution for θ. The first panel of Figure 2 shows the $L(\theta|\tilde{u})$ that corresponds to $\tilde{u} = (1,0)$ for the two items in Figure 1; the second panel is the $p(\theta|\tilde{u})$ that results from $L(\theta|\tilde{u})$ when $p(\theta) = N(0,1)$.

[Figure 2 about here]

2.2 Notation for Missing Responses

Suppose that an examinee provides responses to only a subset of the items. The data thus consist of (i) the identification of the items to which responses are observed and (ii) the responses to those items. We consider inference about θ from this extended observation, assuming the IRT model and item parameters are known, adapting notation and terminology from Little and Rubin (1987) and Rubin (1976):

- $U = (U_1,...,U_n)$ is the (possibly hypothetical) random vector of responses to all items.

- $M = (M_1,...,M_n)$ is a "missing-data indicator," with each element taking a value of 0 or 1. If $m_j=1$, the value of U_j is observed; if $m_j=0$, it is missing.

- $V = (V_1,...,V_n)$ conveys the data that are actually observed: $V_j = U_j$ if $m_j = 1$ but $V_j = *$ if $m_j = 0$, where $*$ indicates that the value of U_j is missing.

An realized value of M, say $\tilde{m}$, effects a partition of U, u, V, and v according to which elements are observed and which are missing. We write $U = (U_{\text{mis}}, U_{\text{obs}})$ to distinguish the missing and observed elements of U; similarly, $u = (u_{\text{mis}}, u_{\text{obs}})$. As with $\tilde{u}$ and $\tilde{m}$, $\tilde{v}$ denotes a realized value of V. Note that v is inferentially equivalent to (u_{obs}, m); and that by (2.1), $f_\theta(u) = f_\theta(u_{\text{mis}}) f_\theta(u_{\text{obs}})$.

Inferences about θ must be based on the data that are actually observed, or $V \equiv (U_{\text{obs}}, M)$. Modeling the hypothetical complete data vector (U, M), even if there is no intention of observing a response to every item, forces us to explicate our beliefs about the relationships among ability, item response, and missingness. To this end, define $g_\phi(m|u) \equiv \text{Prob}(M = m | U = u, \phi)$, with ϕ the possibly vector-valued parameter of the missingness process (which may include θ itself). As we shall see, the form of g will depend on the process, and elements of ϕ can characterize the examinee, the testing situation, or both. So defined,

$$\text{Prob}(U = u, M = m | \theta, \phi) = \text{Prob}(U = u | \theta, \phi) \text{Prob}(M = m | U = u, \theta, \phi)$$
$$= \text{Prob}(U = u | \theta) \text{Prob}(M = m | U = u, \theta, \phi)$$
$$= f_\theta(u) g_\phi(m|u).$$

Whenever all potential responses may not be observed for any reason, even if they all do turn out to be observed, the data are v. The likelihood function is obtained as

$$L(\theta, \phi | \tilde{v}) \equiv L(\theta, \phi | \tilde{m}, \tilde{u}_{\text{obs}}) = \delta_{\theta\phi} \int f_\theta(u_{\text{mis}}, \tilde{u}_{\text{obs}}) g_\phi(\tilde{m} | u_{\text{mis}}, \tilde{u}_{\text{obs}}) du_{\text{mis}}, \qquad (2.3)$$

with $\delta_{\theta\phi}$ taking the value 1 if a value (θ, ϕ) is in the parameter space $\Omega_{\theta\phi}$ and 0 if not. The realized values of observed responses $\tilde{u}_{\text{obs}}$ are constant in (2.3), and marginalization is over the unknown values of the unobserved responses u_{mis}. The observed-data likelihood is thus a weighted average over all complete-data likelihoods for full response vectors u that are in accord with the observed responses to the observed items $\tilde{u}_{\text{obs}}$. The weights are proportional to the probabilities of these potential response patterns for the different values u_{mis}, with $\tilde{m}$ and $\tilde{u}_{\text{obs}}$ fixed. In the context of IRT, the probability for the observed responses can be factored out and brought outside the integral, so that

$$L(\theta, \phi | \tilde{v}) = \delta_{\theta\phi} f_\theta(\tilde{u}_{\text{obs}}) \int f_\theta(u_{\text{mis}}) g_\phi(\tilde{m} | u_{\text{mis}}, \tilde{u}_{\text{obs}}) du_{\text{mis}}. \qquad (2.4)$$

Appropriate likelihood inferences are based on relative values of $L(\theta, \phi | \tilde{v})$ at various values of (θ, ϕ). Bayesian inferences are based on the posterior distribution

$$p(\theta, \phi | \tilde{v}) = K(\tilde{v}) \, L(\theta, \phi | \tilde{v}) p(\theta, \phi), \qquad (2.5)$$

with $p(\theta,\phi)$ the prior distribution for (θ,ϕ). In general, the correct likelihood function for θ under IRT with missing responses involves a nuisance parameter ϕ, and depends not on just the responses that *were* observed, through $f_\theta(\tilde{u}_{\mathrm{obs}})$, but also on the responses that were *not* observed, through $f_\theta(u_{\mathrm{mis}})$ and $g_\phi(\tilde{m}|u_{\mathrm{mis}},\tilde{u}_{\mathrm{obs}})$ under the integral in (2.4).

2.3 Conditions for Ignorability

Ignoring the missingness process when drawing inferences about θ means that instead of using the correct likelihood $L(\theta,\phi|\tilde{v})$, using a facsimile of (2.1) with $\tilde{u}_{\mathrm{obs}}$ alone,

$$L^*(\theta|\tilde{u}_{\mathrm{obs}}) = \delta_\theta f_\theta(\tilde{u}_{\mathrm{obs}}). \tag{2.6}$$

Direct likelihood inferences about θ that ignore the missingness process simply compare values of L^* at various values of θ, and Bayesian inferences that ignore the missingness process proceed from a facsimile of (2.2) obtained as

$$L^*(\theta|\tilde{u}_{\mathrm{obs}})p(\theta). \tag{2.7}$$

It is a pleasant state of affairs when the missingness process can be ignored, since (2.6) and (2.7) don't require the specification of g, and standard computing algorithms can be used. Depending on why the missing responses were missing, however, these procedures need not lead to the correct inferences. Rubin (1976) specifies conditions under which a missingness process can be ignored under sampling distribution, direct likelihood, and Bayesian inference. Useful sufficient conditions for ignorability under direct likelihood and Bayesian inference, the focus of this paper, involve the following concepts:

DEFINITION. Missing responses are **missing completely at random** (MCAR) if for each value of ϕ and for each fixed value m, $g_\phi(m|u)$ takes the same value for all u. That is, $g_\phi(m|u) = g_\phi(m)$.

DEFINITION. Missing responses are **missing at random** (MAR) if for each value of ϕ and for all fixed values m and u_{obs}, $g_\phi(m|u_{\mathrm{mis}},u_{\mathrm{obs}})$ takes the same value for all u_{mis}. That is, $g_\phi(m|u) = g_\phi(m|u_{\mathrm{obs}})$.

DEFINITION. The parameter θ is **distinct** (D) from ϕ if their joint parameter space factors into a θ-space and a ϕ-space, and when prior distributions are specified for θ and ϕ, they are independent.

Remarks. (a) MCAR implies MAR. (b) For direct-likelihood distinctness to be satisfied, conditioning on ϕ must not change the support of the likelihood function for θ. For Bayesian distinctness, conditioning on ϕ must also not change belief about θ. (c) Taken

together, MCAR and D imply that the values of both the observed and the missing responses are independent of the pattern of missingness. MAR and D together imply that the values of the missing responses are independent of the pattern of missingness, conditional on the values of the observed responses.

We are now in a position to summarize Rubin's results for direct-likelihood and Bayesian inferences. First, a more easily verified *sufficient* condition:

THEOREM 2.1 (Rubin, 1976, pp. 581). When making direct-likelihood or Bayesian inferences about θ, it is appropriate to ignore the process that causes missing data if MAR and D are satisfied.

Proof. When MAR is satisfied, g does not depend on u_{mis} and can be brought out of the integral in (2.3), which then simply integrates to one. If D is satisfied as well, $L(\theta, \phi | \tilde{v})$ depends on θ only through $f_\theta(\tilde{u}_{\mathrm{obs}})$. $\qquad \square$

Under weaker conditions, the integral need not drop out but its value does not depend on θ. *Necessary and sufficient* (NS) conditions are given below, without proofs.

THEOREM 2.2 (Rubin, 1976, pp. 586). Suppose $L(\theta, \phi | \tilde{v}) > 0$ for all $\theta \in \Omega_\theta$. All likelihood ratios for θ ignoring the process that causes missing data are correct for all $\phi \in \Omega_\phi$, if and only if (a) $\Omega_{\theta\phi} = \Omega_\theta \times \Omega_\phi$ and (b) for each $\phi \in \Omega_\phi$, the quantity

$$E_{u_{\mathrm{mis}}} \left\{ g_\phi(\tilde{m} | u_{\mathrm{mis}}, \tilde{u}_{\mathrm{obs}}) | \tilde{m}, \tilde{u}_{\mathrm{obs}}, \theta, \phi \right\} \qquad (2.8)$$

takes the same positive value for all θ.

Remarks. Theorem 2.2 says that for direct-likelihood ignorability to hold in the IRT context, given any value of the missingness parameter and the observed data $\tilde{v}$, the probability of the observed pattern of missingness must be the same for all values of θ. This is true if MAR and D hold, since these constitute sufficient conditions for ignorability. If D holds but MAR does not, the varying values of $g_\phi(\tilde{m} | u_{\mathrm{mis}}, \tilde{u}_{\mathrm{obs}})$ under the integral in (2.3) must be exactly counterbalanced by varying values of $f_\theta(u_{\mathrm{mis}}, \tilde{u}_{\mathrm{obs}})$. While it is straightforward to construct artificial examples in which this happens, it appears rare in practice to find applications in which the conditions in Theorem 2.2 are satisfied but MAR is not. Section 7 below, for example, discusses the counter-intuitive circumstances that would have to hold if intentional omitting were to meet this condition.

THEOREM 2.3 (Rubin, 1976, pp. 587). The posterior distribution of θ ignoring the process that causes missing data equals the correct posterior distribution of θ if and only if

$$E_{u_{\text{mis}},\phi}\left\{g_\phi\left(\tilde{m}|u_{\text{mis}},\tilde{u}_{\text{obs}}\right)\middle|\tilde{m},\tilde{u}_{\text{obs}},\theta\right\} \tag{2.9}$$

takes a constant positive value.

Remark. Theorem 2.3 says that for Bayesian ignorability to hold in the IRT context, then given the observed data $\tilde{v}$ the probability of the observed pattern of missingness must be the same for all values of θ.

3.0   Alternate Test Forms

"Alternate test forms" are sets of items that all fit the same IRT model, and the test administrator is indifferent as to which form an examinee is presented. The item sets on different forms may overlap. The form an examinee receives depends on a random process specified by the administrator, such as a coin flip or a form-spiraling scheme. In practice, IRT inferences about θ from alternate test forms are commonly based on $L^*\left(\theta|\tilde{u}_{\text{obs}}\right)$.

The use of K alternate test forms implies that only K missingness patterns, say $\left\{m^{(1)},\ldots,m^{(k)},\ldots,m^{(K)}\right\}$, can occur, where all the item-level elements of pattern $m^{(k)} \equiv \left(m_1^{(k)},\ldots,m_n^{(k)}\right)$ are zero except those which correspond to the items that appear in Form k. Denote by ϕ_k the administrator-determined values $\text{Prob}(M=m^{(k)})$. Assuming an LI (locally independent) IRT model means that $f_\theta(u)$ is as given in (2.1); that is, the values of item responses are governed by θ alone, regardless of which items would be administered. Even though the items of only one form will actually be presented, it is possible to express our assumptions about the connection between the (hypothetical) values of the complete response pattern and the probability of the missingness pattern as follows:

$$g_\phi(m|u) = \begin{cases} \phi_k & \text{for all } u \text{ if } m = m^{(k)} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

THEOREM 3.1. Random assignment of alternate test forms satisfies MCAR, and therefore MAR as well.

Proof. This follows immediately from (3.1), since the values of g do not depend on u. $\square$

THEOREM 3.2. The missingness induced by random administration of alternate test forms is ignorable under direct likelihood and Bayesian inference.

Proof. By Theorem 3.1, MAR is satisfied. Verifying D for likelihood inference requires that the θ and ϕ parameter spaces are distinct; this follows from (3.1). Verifying D for Bayesian inference requires that prior beliefs about θ and ϕ be independent; this also follows from (3.1). Ignorability follows from Theorem 2.1. $\square$

4.0 Targeted Testing

"Targeted testing" involves multiple test forms in which the distributions of item difficulty differ purposefully from form to form. Exploiting the facts that (1) estimates of θ are more precise when an examinee is administered items with difficulties near the value of θ, and (2) covariates y that are related to θ may be available, targeted testing uses an examinee's covariate to select a test form that is likely to be more informative than other otherwise similar forms. For example, an easy form and a hard form might be constructed from a set of n items calibrated together under the same IRT model, and the easy form administered to first graders and the hard form to second graders.

As with alternate test forms, the existence of K forms for targeted testing implies that only K patterns of M, again denoted $\left\{m^{(1)},...,m^{(K)}\right\}$, can be realized. As with alternate test forms, the missingness parameter ϕ is the vector of indicator variables ϕ_k indicating the test form selected for the examinee. The parameter of the missingness process now consists of the administrator-determined values $\phi_k(y) \equiv \text{Prob}\left(M = m^{(k)}|y\right)$, which indicate the probability that an examinee with covariate y will be administered Form k. For at least one k and two values y' and y'', $\phi_k(y') \neq \phi_k(y'')$. This happens when the test administrator knows $p(\theta|y') \neq p(\theta|y'')$ and that the difficulty of Form k is better suited to the typical examinee with $Y=y'$ than one with $Y=y''$, or vice versa.

THEOREM 4.1. Targeted assignment of test forms based on examinee covariates satisfies MCAR, and therefore MAR as well.

Proof. Values of g do not depend on u. $\square$

THEOREM 4.2. The missingness induced by administration of test forms based on examinee covariates y is ignorable under direct likelihood inference if all values of θ can occur at all values of y.

Proof. By Theorem 4.1, MAR is satisfied. Verifying D for likelihood inference requires that θ is distinct from ϕ; that is, conditioning on ϕ must not change the support of the likelihood function for θ. This condition fails under targeted testing only if, for some assignments of test forms, certain values of θ cannot occur. $\square$

THEOREM 4.3. The missingness induced by administration of test forms based on examinee covariates y is ignorable under Bayesian inference only if θ and y are independent (i.e., the targeting is wholly ineffective).

Proof. Under the targeted-testing missingness mechanism, g does not depend on u and the missingness parameter ϕ is fixed a priori by the test administrator, so (2.9) simplifies to $\left\{ g_\phi(\tilde{m}) \big| \tilde{m}, \theta \right\}$, or the probability of a given missingness pattern given θ. By Theorem 2.3, ignorability obtains under Bayesian inference iff this quantity takes a constant positive value. But since targeted testing means that g is not constant with respect to y, g will be constant for θ only if y is independent of θ. $\square$

THEOREM 4.4. The missingness induced by targeted testing is ignorable under Bayesian inference conditional on y; i.e., the correct posterior is proportional to $L^*(\theta | \tilde{u}_{obs}) p(\theta | y)$.

Proof. As in the preceding proof, under targeted testing the expression (2.9) simplifies because g does not depend on u and the missingness parameter ϕ is fixed a priori by the test administrator. If we condition on y, it becomes $\left\{ g_\phi(\tilde{m}) \big| \tilde{m}, \theta, y \right\}$, or, by the definition of g, simply $\left\{ g_\phi(\tilde{m}) \big| \tilde{m}, y \right\}$—a constant with respect to θ, as required by Theorem 2.3. $\square$

Remarks. By Theorem 4.2, the correct value of the MLE is obtained for θ under targeted testing. For correct Bayesian inference, however, the relationship between y and θ must be taken into account, so $L^*(\theta | \tilde{u}_{obs}) p(\theta | y)$ yields correct inferences but $L^*(\theta | \tilde{u}_{obs}) p(\theta)$ generally does not.

5.0 Adaptive Testing

As noted above, IRT measurement is more precise if an examinee is administered items that are informative in the neighborhood of the value of θ (Wainer et al., 1990). Adaptive testing uses an examinee's preceding responses to select each next item to administer. Under the Rasch model, for example, an examinee answering items correctly

would be administered successively more difficult items, and an examinee answering incorrectly would be administered successively easier items.

The datum observed in adaptive testing is a sequence of $N \leq n$ ordered pairs, $S = \left(\left(I_1, U_{obs(1)} \right), \ldots, \left(I_N, U_{obs(N)} \right) \right)$, where I_k identifies the k'th item administered and $U_{obs(k)}$ is the response to that item. Define the partial response sequence S_k as the first k ordered pairs in S, with the null sequence s_0 representing the status as the test begins. Testing may continue until a desired level of precision is reached, a predetermined number of items has been administered, or a specified number of correct or of incorrect responses has been observed. Augment the collection of items with the fictitious Item 0, the selection of which corresponds to a decision to terminate testing. It can be written as the N+1$^{\text{st}}$ item in the adaptive test, but no response is associated with it.

A test administrator defines an adaptive test design by specifying, for all items j and all realizable partial response sequences s_k, the probabilities $\phi(j, s_k)$ that Item j will be selected as the k+1$^{\text{st}}$ test item, after the partial response sequence s_k has been observed from an examinee. Under Bayesian minimum variance item selection, for example, the as-yet-unadministered item that minimizes the expected posterior variance of θ with respect to the current distribution $p(\theta|s_k)$ is chosen as the k+1$^{\text{st}}$ item with probability one (Owen, 1975). Note that the value of s conveys the value of v, because m_j=1 if i_k=j for some $k \in (1, \ldots, N)$ and m_j=0 if not, and the responses to the administered items constitute u_{obs}.

THEOREM 5.1. The conditional distribution of response sequence S is given by

$$\text{Prob}(S = s|\theta) = f_\theta\left(u_{\text{obs}}\right) \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right). \tag{5.1}$$

Proof. The probability of S for an examinee with ability θ can be constructed sequentially. The probability of selection for the first item is $\phi(i_1, s_0)$. The probability of response u_{i_1} to Item i_1 is given by the IRT model as $f_\theta\left(u_{i_1}\right)$, which does not depend on the fact that Item i_1 happened to have been presented first. The probability of selection for the second item *given* s_1 is $\phi(i_2, s_1)$, which depends on the value of u_{i_1} but not on θ given u_{i_1}. The probability of the corresponding response is $f_\theta\left(u_{i_2}\right)$, again independent of the identification of, and the response to, the first item. Continuing in this manner through the decision to stop testing, or the selection of Item 0 as the N+1$^{\text{st}}$ item, yields

$$\text{Prob}(S = s | \theta) = \prod_{k=1}^{N} f_\theta\left(u_{i_k}\right) \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right)$$

$$= f_\theta\left(u_{\text{obs}}\right) \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right).$$

(5.2)

□

THEOREM 5.2. *The missingness mechanism induced by adaptive testing satisfies MAR.*

Proof. Since S conveys the values of m and u_{obs}, the probability of (m, u) is the product of the probability of (i) observing a response sequence that implies m and u_{obs}, and (ii) the probability of u_{mis}. Denote by $T_m = \left\{s : M = m \text{ and } U_{\text{obs}} = u_{\text{obs}}\right\}$ the set of response sequences with missingness patterns and observed item responses that match those of (m, u). Then

$$\text{Prob}(M = m, U = u) = \text{Prob}\left(S \in T_m\right) \times \text{Prob}\left(U_{\text{mis}} = u_{\text{mis}}\right)$$

$$= \int \text{Prob}\left(S \in T_m | \theta\right) \times \text{Prob}\left(U_{\text{mis}} = u_{\text{mis}} | \theta\right) \partial\theta$$

$$= \int \left\{ \sum_{s \in T_m} f_\theta\left(u_{\text{obs}}\right) \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right) \right\} f_\theta\left(u_{\text{mis}}\right) \partial\theta$$

$$= \left\{ \sum_{s \in T_m} \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right) \right\} \left\{ \int f_\theta\left(u_{\text{obs}}\right) f_\theta\left(u_{\text{mis}}\right) \partial\theta \right\}$$

$$= \left\{ \sum_{s \in T_m} \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right) \right\} \times \text{Prob}(U = u).$$

Since $g_\phi(m|u) = \text{Prob}(M = m, U = u) / \text{Prob}(U = u)$,

$$g_\phi(m|u) = \sum_{s \in T_m} \prod_{k=1}^{N+1} \phi\left(i_k, s_{k-1}\right),$$

(5.3)

which does not depend on u_{mis}, as required to satisfy MAR.

□

THEOREM 5.3. *The missingness induced by adaptive testing is ignorable under direct likelihood and Bayesian inference.*

Proof. By Theorem 5.1, the conditional probability of the observation S given θ factors into two terms, namely $f_\theta\left(u_{\text{obs}}\right)$ and a term that does not depend on θ. The term $L(\theta, \phi | \tilde{v})$ in direct likelihoods and Bayesian posteriors thus reduces to $L^*\left(\theta | \tilde{u}_{\text{obs}}\right)$, as required for ignorability.

□

Alternative Proof. The missingness mechanism g in adaptive testing is specified a priori, and does not depend on θ; thus, D is satisfied in both the likelihood and Bayesian senses. By Theorem 5.2, the missingness mechanism is also MAR. The sufficient conditions in Theorem 2.1 for ignorability are satisfied. $\square$

Remark. Ignorability under direct likelihood inference means that $L^*\left(\theta | \tilde{u}_{\text{obs}}\right)$ yields the correct value of the MLE $\hat{\theta}$ from an observed s, but it does not justify the sampling-distribution interpretation of $\hat{\theta}$; that is, the correct point estimate is identified but no claims about its distribution in repeated samples for fixed θ necessarily follow. It can be shown that Rubin's (1976) NS conditions for ignorability under sampling-distribution inference would require that the probability of any given missingness pattern be the same no matter what values the responses took (Mislevy & Wu, 1988). But since by definition adaptive tests produce missingness patterns as a function of the response values that are observed, only a degenerate adaptive testing scheme would satisfy this condition. Concluding that the item selection mechanism is not ignorable for sampling distribution inference in general means that the correct sampling distribution for $\hat{\theta}$ must be verified with respect to repeated administrations of the entire adaptive test. Chang and Ying (1996) consider the sampling variance of $\hat{\theta}$ to the second order derivative of L^*, and offer some large-sample conditions under which the latter is a reasonable large-sample approximation of the former.

6.0   Not-Reached Items

IRT is intended for "power" tests, or those in which an examinee's chances of responding correctly would not differ appreciably if the time limit were more generous. Time limits are typically chosen to allow most examinees to respond to all items, but some examinees don't have time to answer all of them (see Yamamoto & Everson, 1995, for analyses of the situation in which examinees respond in accordance with an IRT model until time is nearly up, then switch to random responding.) This section concerns the items that an examinee does not reach, assuming the following conditions:

(i)   An LI IRT model would give response probabilities if the examinee had interacted meaningfully with all the items.

(ii)   The examinee has no information about the difficulty or content of the items at the end of the test, but has decided to work instead from the beginning of the test toward the end, answering all items along the way, until time expires.

(iii)   All n items are administered. (If they are not, the results of this section must be combined with those of Sections 3 or 4, as appropriate.)

THEOREM 6.1. Under conditions (i)-(iii), the missingness induced by failing to reach the end of the test because of time limitations satisfies MCAR, and therefore MAR as well.

Proof. Under conditions (i)-(iii), $n+1$ patterns of missingness can occur: For $k = 0,...,n$, let $m^{(k)}$ denote the string of n-k 1's followed by k 0's. That is, $m^{(k)}$ is the missingness pattern of an examinee that has not reached the last k items. The missingness process is characterized by the examinee speed parameter $\phi = (\phi_0,...,\phi_n)$, where $\phi_k =$ Prob$(M = m^{(k)})$, the probability that the examinee will not reach the last k items, and

$$g_\phi(m|u) = \begin{cases} \phi_k & \text{for all } u \text{ if } m = m^{(k)} \\ 0 & \text{otherwise.} \end{cases}$$

As required for MCAR, g does not depend on u.   □

THEOREM 6.2. Under conditions (i)-(iii), the missingness induced by failing to reach the end of the test because of time limitations is ignorable under direct likelihood inference if all values of θ can occur at all values of ϕ.

Proof. The requirement that all values of θ can occur at all values of ϕ is D, as it pertains to direct likelihood inference. By Theorem 6.1, the missingness is MAR under conditions (i)-(iii). By Theorem 2.1, D and MAR give ignorability under direct likelihood inference.□

THEOREM 6.3. Under conditions (i)-(iii), the missingness induced by failing to reach the end of the test because of time limitations is ignorable under Bayesian inference if $p(\theta,\phi)$ $= p(\theta)p(\phi)$.

Proof. The requirement that $p(\theta,\phi) = p(\theta)p(\phi)$ is D, as it pertains to Bayesian inference. Under conditions (i)-(iii), the missingness is MAR. By Theorem 2.1, these two conditions imply ignorability under Bayesian inference.   □

THEOREM 6.4. Under conditions (i)-(iii), the missingness induced by failing to reach the end of the test because of time limitations is ignorable under Bayesian inference only if, for each $k = 0,...,n$, the expected value of $\phi^{(k)}$ is constant across all values of θ.

Proof. For a given not-reached pattern $m^{(k)}$, Equation (2.9), the NS condition for ignorability under Bayesian inference simplifies to $\int \phi^{(k)} p(\phi^{(k)}|\theta) \partial\phi^{(k)}$ because g does not depend on u.   □

Remark. Since independence of θ and ϕ implies, but is not implied by, constant conditional means for ϕ, Theorem 6.4 gives a weaker condition than Theorem 6.3 for ignorability of not-reached items under Bayesian inference.

By Theorem 6.2, if conditions (i)-(iii) hold then not-reached responses are ignorable under direct likelihood inference; in particular, the correct value is obtained for the MLE. For ignorability to hold under Bayesian inference, however, it is further necessary that the expected value of each speededness subparameter is constant across all values of θ (Theorem 6.4). Empirical evidence suggests that this is not generally true. Van den Wollenberg (1979), for example, reports significant positive correlations between percent-correct scores on the first eleven items (which were reached by all examinees) and the total number of items reached, in four of six intelligence tests in the ISI battery (Snijders, Souren, and Welten, 1963). Bayesian inference about θ would take this relationship into account by using the correct posterior distribution, or

$$p(\theta,\phi|\tilde{v}) \propto p(\tilde{v}|\theta,\phi)p(\theta,\phi)$$

$$= \left[\int p(\tilde{m},\tilde{u}|\theta,\phi)\partial u_{\mathrm{mis}}\right]p(\theta,\phi)$$

$$= \left[\int g_\phi(\tilde{m}|u_{\mathrm{mis}},\tilde{u}_{\mathrm{obs}})f_\theta(u_{\mathrm{mis}},\tilde{u}_{\mathrm{obs}})\partial u_{\mathrm{mis}}\right]p(\theta,\phi)$$

$$= f_\theta(\tilde{u}_{\mathrm{obs}})g_\phi(\tilde{m})p(\theta,\phi) \qquad \text{[by MCAR]}$$

$$= L^*(\theta|\tilde{u}_{\mathrm{obs}})g_\phi(\tilde{m})p(\theta,\phi).$$

Further, the correct marginal posterior for θ is obtained as

$$p(\theta|\tilde{v}) = \int L^*(\theta|\tilde{u}_{\mathrm{obs}})g_\phi(\tilde{m})p(\theta,\phi)\partial\phi$$

$$= L^*(\theta|\tilde{u}_{\mathrm{obs}})\left[\int g_\phi(\tilde{m})p(\phi|\theta)\partial\phi\right]p(\theta)$$

$$= L^*(\theta|\tilde{u}_{\mathrm{obs}})\left[\int p(\tilde{m}|\phi)p(\phi|\theta)\partial\phi\right]p(\theta)$$

$$= L^*(\theta|\tilde{u}_{\mathrm{obs}})p(\tilde{m}|\theta)p(\theta)$$

$$\propto L^*(\theta|\tilde{u}_{\mathrm{obs}})p(\theta|\tilde{m}).$$

7.0   Intentionally Omitted Items

A missing response is an intentional omission when an examinee is administered the item, has time to appraise it, and decides for whatever reason not to respond. After arguing that such omissions can't generally be considered ignorable, we discuss ways to deal with them. Several solutions suggested in the test theory literature and an approach suggested by the present analyses are considered.

7.1 Omitting Behavior and Ignorability

Used with tests of dichotomous items, a test score $T(v)$ summarizes a pattern of rights, wrongs, and omits for the purposes of comparing, selecting, or describing examinees. Formula scores take the form

$$T(v) = R(v) - C \cdot W(v), \qquad (7.1)$$

where $R(v)$ and $W(v)$ are counts of right and wrong responses and C, with $0 \le C \le 1$, is a constant selected by the test administrator. Setting $C = 0$ gives number-right scores; $C = 1$ gives right-minus-wrong scores; and for multiple choice items with A alternatives, $C = 1/(A-1)$ gives the familiar "corrected-for-guessing" scores.

THEOREM 7.1.1. For a set of items for which a SMURFLI[2] IRT model holds, $E\{T(U)|\theta\}$ is increasing in θ.

Proof.

$$E\{T(U)|\theta\} = E\{R(U) - C \cdot W(U)|\theta\}$$

$$= \sum_{j=1}^{n} \mathrm{Prob}(U_j = 1|\theta) - C \sum_{j=1}^{n} [1 - \mathrm{Prob}(U_j = 1|\theta)]$$

$$= (1 + C) \sum_{j=1}^{n} \mathrm{Prob}(U_j = 1|\theta) - Cn.$$

By monotonicity, $\theta' > \theta'' \Rightarrow \mathrm{Prob}(U_j = 1|\theta') > \mathrm{Prob}(U_j = 1|\theta'')$ for all items j, so the final equation implies that $\theta' > \theta'' \Rightarrow E\{T(U)|\theta'\} > E\{T(U)|\theta''\}$. □

THEOREM 7.1.2. For any partitioning of items inducing $U = (U', U'')$, $E\{T(U)\} = E\{T(U')\} + E\{T(U'')\}$. Similarly, $E\{T(V)\} = E\{T(V')\} + E\{T(V'')\}$, $T(u) = T(u') + T(u'')$, and $T(v) = T(v') + T(v'')$.

Proof. These results follow from the linearity of the definition of T. □

Since a correct response to an item gives a higher test score than an incorrect response, examinees who wants to obtain high scores will make responses they think are correct. How examinees will respond to an item about which they are unsure depends at least partly on how the test will be scored (Sabers and Feldt, 1968). They maximize their expected scores by answering items for which they think their probability of being correct is at least $C/(1 + C)$. Under number-right scoring, they should answer every item; under corrected-for-guessing scoring, they should answer those which they think the probability

is at least $1/A$; and under right-minus-wrong scoring, they should answer those which they thinks the probability is at least $1/2$ ($C/(1+C)$ when $C=1$). Examinees may differ in the accuracy of their estimates, their confidence about them, and their propensities to omit rather than make responses about which they are uncertain. Such characteristics of an examinee, as they are evoked under given test administration conditions, constitute the missingness parameter ϕ in the case of intentional omission. For example, analyzing responses to items that examinees originally omitted under right-minus-wrong scoring, Sherriffs and Boomer (1954) did find that about half of the omitted responses would have been correct among examinees who scored low on a risk-aversion scale, but nearly two-thirds would have been correct among examinees with high risk-aversion scores. Intuition and empirical evidence (e.g., Stocking, Eignor, & Cook, 1988) suggest that the following three conditions typify omitting behavior in educational testing:

(iv)   For any given θ and a given item, examinees are more likely to omit items when they think their answers would be incorrect than items they think their answers would be correct.

(v)   As θ increases, an examinee is more likely to recognize when a response would be *correct*, and are thus *less* likely to omit it; i.e., for all items j,

$$\theta'' > \theta' \Rightarrow g_{\phi,\theta''}\left(M_j = 0 | U_j = 1\right) < g_{\phi,\theta'}\left(M_j = 0 | U_j = 1\right),$$

where θ has been made explicit as a subscript of g to emphasize the dependence of omitting behavior on θ.

(vi)   Similarly, as θ increases, an examinee is more likely to recognize when a response would be *incorrect*, and are thus *more* likely to omit it; i.e., for all items j,

$$\theta'' > \theta' \Rightarrow g_{\phi,\theta''}\left(M_j = 0 | U_j = 0\right) > g_{\phi,\theta'}\left(M_j = 0 | U_j = 0\right).$$

THEOREM 7.1.3. For a given $\tilde{u}_{\text{obs}}$ and an $\tilde{m}$ in which missing responses are due to intentional omission, the missingness process is MAR only if this missingness pattern is equally likely for all values of u_{mis}.

Proof. This follows immediately from the definition of MAR.   □

Remarks. (a) If Condition (iv) above holds, examinees are more likely to omit items when they think their responses would be wrong rather than right. MAR would imply that their perceived probabilities of correctness are independent of the probabilities given by the IRT model; that is, Conditions (v) and (vi) could not also hold. (b) MAR (along with D) is merely sufficient for ignorability, and ignorability can hold when MAR does not. If

ignorability does hold, though, the following counter-intuitive condition holds for expected score of the omitted responses.

THEOREM 7.1.4. Suppose a SMURFLI[(2)] IRT model holds in a domain of items, and the missingness induced by intentional omission is ignorable under likelihood or Bayesian inference. For any given pattern of omits $(\tilde{m})$ and responses to observed items $(\tilde{u}_{\text{obs}})$, the probability of this missingness pattern is (a) constant with respect to θ, even though (b) the expected score of the missing responses $T(U_{\text{mis}})$ is increasing with respect to θ.

Proof. Regarding (a): The NS condition for ignorability under likelihood inference (2.8) requires that for each value of ϕ, the value of $E_{u_{\text{mis}}}\left\{g_\phi(\tilde{m}|u_{\text{mis}},\tilde{u}_{\text{obs}})|\tilde{m},\tilde{u}_{\text{obs}},\theta,\phi\right\}$ (the probability that missingness pattern $\tilde{m}$ will be observed given $\tilde{u}_{\text{obs}}$, ϕ, and θ) is constant with respect to θ. Regarding (b): By local independence, the distribution of U_{mis} (given $\tilde{m}$) depends on θ, and not on ϕ or $\tilde{u}_{\text{obs}}$. By Theorem 7.1.1, $E\{T(U_{\text{mis}})|\theta\}$, is increasing in θ. The same argument holds for ignorability under Bayesian inference, since, by Theorem 2.3, it is required that the expectation of (2.8) over ϕ, or (2.9),is constant with respect to θ. □

Remark. Theorem 7.1.4 says that if ignorability holds under a SMURFLI[(2)] IRT model, high-ability examinees are just as likely to produce any given missingness pattern as low-ability examinees (given $\tilde{u}_{\text{obs}}$), even though the missing responses have a greater expected contribution to their total test score. This result belies (iv)-(vi), since higher-ability examinees are more likely to make correct responses to the missing items, more likely to recognize they are correct, and more likely to make the responses rather than omit them.

COROLLARY 7.1.5. Suppose a SMURFLI[(2)] IRT model holds for a one-item domain, and the missingness induced by intentional omission is ignorable under likelihood or Bayesian inference. The probability that the response will be omitted is constant with respect to θ even though the probability that it will be answered correctly increases with θ.

7.2 Modeling Intentional Omissions

Since ignorability is not generally satisfied for direct likelihood inference, it is necessary to base inference on $p(v|\theta,\phi)$ when omitting is a possibility. Accepting that omits are not ignorable means $g_\phi(m|u)$ depends on u_{mis}, and cannot be determined from observations of V alone. Implementing any approach for modeling intentional omission thus requires that the analyst must either specify the mechanism of the omitting process a

priori, or estimate it from an experiment with the same items and similar examinees in which the values of item responses that were originally omitted are subsequently obtained. Section 7.2.1 presents an approach from the perspective of Rubin's model, and Sections 7.2.2-7.2.4 evaluate various alternative approaches that have appeared in the literature.

7.2.1 An Approach Based on Rubin's Model

Rubin's framework begins with a full model for response and omission, (U,M), using the form $\text{Prob}(U = u, M = m|\theta, \phi) = g_\phi(m|u) f_\theta(u)$ with $f_\theta(u)$ an LI IRT model. This section sketches some specifics about how this approach might be implemented, bearing in mind that $g_\phi(m|u)$ cannot be estimated from observations of V alone because its values depends on u_{mis}. The main ideas are (i) viewing the missingness parameter ϕ as the concatenation of item-specific missingness parameters η_j and examinee-specific parameters ω, and (ii) assuming conditional independence across the expanded item response (u_j, m_j) given the extended examinee parameter (θ, ω) and extended item parameters (β_j, η_j). Specifically, it may be posited that

$$\text{Prob}(U = u, M = m|\theta, \phi) = \prod_{j=1}^{n} g_\phi(m_j|u_j) f_\theta(u_j),\qquad(7.2)$$

where

$$g_\phi(m_j|u_j) = \text{Prob}(M_j = m_j|U_j = u_j, \theta, \omega, \eta_j).\qquad(7.3)$$

For dichotomous items, one plausible form for (7.3) is a pair of linear logistic regression functions for the probability of omitting incorrect and correct responses, given an items' tendency to provoke omission (η_j), the examinee's ability (θ), and the examinee's tendency to omit (ω):

$$\text{Prob}(M_j = 0|U_j = u_j, \theta, \omega, \eta_j) = \begin{cases} \Psi(\eta_{j01} + \eta_{j02}\theta + \omega) & \text{if } u_j = 0 \\ \Psi(\eta_{j11} + \eta_{j12}\theta + \omega) & \text{if } u_j = 1, \end{cases}$$

where $\Psi(z) = \exp(z)/[1 + \exp(z)]$. More ambitious models would allow for the dependence of g on covariates y as well, since empirical evidence suggests that omitting behavior can vary systematically with factors such as gender and culture (e.g., Wolf, 1977, p. 33). With g sufficiently specified in this or some other manner, and in conjunction with an IRT model, likelihood and Bayesian inference about θ from v can proceed from (2.4) or (2.5) respectively.

7.2.2 Filling in the Blanks

Lord (1974) suggested that omits on dichotomously-scored multiple-choice items under guessing-corrected scoring can be handled with standard IRT estimation routines if they are treated as fractionally correct, with the value $c=1/(\text{\# alternatives})$. He assumed "rational" omitting behavior: Examinees omit items only if their chances of responding correctly would have been c, so $\text{Prob}\left(U_j = 1 \middle| M_j = 0\right) = c$ for all items and all θ. Now the log likelihood to be maximized to obtain $\hat{\theta}$ if there is no possibility of missing responses is

$$\ell(\theta|\tilde{u}) = \delta_\theta \sum_{j=1}^n \left[\tilde{u}_j \log P_j(\theta) + \left(1 - \tilde{u}_j\right)\log Q_j(\theta)\right], \tag{7.4}$$

where $P_j(\theta) = \text{Prob}\left(U_j = 1\middle|\theta, \beta_j\right)$ and $Q_j(\theta) = 1 - P_j(\theta)$. Lord proposed maximizing the pseudo log likelihood given by

$$\ell^{**}(\theta|\tilde{v}) = \delta_\theta \sum_{j=1}^n \left[w_j \log P_j(\theta) + \left(1 - w_j\right)\log Q_j(\theta)\right], \tag{7.5}$$

where $w_j = \tilde{u}_j$ if $m_j = 1$ and $w_j = c$ if $m_j = 0$.

LEMMA 7.2.1. The likelihood based on the hypothetical complete data can be written as the product of two factors, one of which involves only u and θ, and not the missingness process or m.

Proof. By definitions and elementary properties of probability,

$$\text{Prob}\left(U = u, M = m\middle|\theta, \phi\right) = \text{Prob}\left(M = m\middle|U = u, \theta, \phi\right)\text{Prob}\left(U = u\middle|\theta, \phi\right) = g_\phi(m|u)f_\theta(u).$$

Viewed as a likelihood,

$$L(\theta, \phi|u, m) = \delta_{\theta\phi}L_1(\theta, \phi|u, m) \times \delta_\theta L_2(\theta|u),$$

with $L_1(\theta, \phi|u, m) = g_\phi(m|u)$ and $L_2(\theta|u) = f_\theta(u)$. $\qquad\square$

THEOREM 7.2.2. Consider a domain of multiple-choice items for which a SMURFLI[2] IRT model holds, but for which examinees may intentionally omit responses. If $\text{Prob}\left(U_j = 1\middle|M_j = 0\right) = c$ for all items and all θ, then Lord's (1974) pseudo log likelihood (7.5) is the expectation of the log of a conditional likelihood function for θ; specifically,

$$\ell^{**}(\theta|\tilde{v}) = \delta_\theta E_{u_{\text{mis}}}\left\{\log L_2(\theta|u)\middle|\tilde{u}_{\text{obs}}\right\},$$

where $L_2(\theta|u)$ is as given in Lemma 7.2.1.

Proof.

$$E_{u_{mis}}\left\{\log L_2(\theta|u)|\tilde{u}_{obs}\right\}$$

$$= E_{u_{mis}}\left\{\sum_{j:m_j=1}\left[\tilde{u}_j\log P_j(\theta)+\left(1-\tilde{u}_j\right)\log Q_j(\theta)\right]+\sum_{j:m_j=0}\left[u_j\log P_j(\theta)+\left(1-u_j\right)\log Q_j(\theta)\right]\right\}$$

$$= \sum_{j:m_j=1}\left[\tilde{u}_j\log P_j(\theta)+\left(1-\tilde{u}_j\right)\log Q_j(\theta)\right]+E_{u_{mis}}\left\{\sum_{j:m_j=0}\left[u_j\log P_j(\theta)+\left(1-u_j\right)\log Q_j(\theta)\right]\right\}$$

$$= \sum_{j:m_j=1}\left[\tilde{u}_j\log P_j(\theta)+\left(1-\tilde{u}_j\right)\log Q_j(\theta)\right]+\sum_{j:m_j=0}\left[c\log P_j(\theta)+(1-c)\log Q_j(\theta)\right].$$

This final expression, multiplied by δ_θ, is Lord's $\ell^{**}(\theta|\tilde{v})$. $\square$

Remarks. (a) It is a standard technique in likelihood estimation to eliminate nuisance variables from a problem by factoring the likelihood, and basing inference on only those factors which do not involve the nuisance variable. Doing so yields "limited information" inferences, so called because they forego information about the target parameter contained in the neglected factors. In this case, Lord's solution does not use information about θ conveyed by m through L_1. (b) Maximizing the expected value of the log likelihood of L_2 with respect to the missing responses is an instance of the general approach to inference with missing data described in Dempster, Laird, and Rubin (1977); specifically, it is a one step "EM" solution. (c) Taken together, (a) and (b) justify Lord's (1974) maximization of (7.5) as an "expected limited-information" MLE for θ.

The foregoing analysis yields insight into other treatments of omits that impute values for u_{mis}. Supplying random responses that are correct with probability c provides a crude numerical approximation of (7.5), leading to a maximizing value which has the same expectation as when the integration is carried out in closed form in the proof of Theorem 7.2.2. This practice is justified by the same assumptions as Lord's (1974) approach. Supplying incorrect responses for omits also leads to an "expected limited-information" MLE for θ, but under the assumption that responses to omitted items would surely have been incorrect; that is, $\text{Prob}\left(U_j=1|M_j=0\right)=0$ for all items and all θ. This is implausible for multiple-choice items for which even the least able examinees have nontrivial probabilities of success through guessing, so in this case supplying incorrect responses for omits biases inference about θ downward.

Lord addressed "rational" omitting behavior, in that the expectation of correctness for an omitted response is always c, the value associated with the optimal omitting strategy. As noted, however, studies of responses to items originally omitted show that not all examinees behave in this manner. The tendency to omit when probabilities of success may be higher than c and can be associated with personality characteristics, demographic variables, and level of ability. This approach biases estimates of θ downward for risk-aversive examinees. Although Section 7.2.1 showed how such dependencies can be taken into account, it is by no means certain that this should be done; to do so effectively adjusts scores upward or downward in accordance with examinee background characteristics, which may be objectionable on the grounds of fairness. Assuming rational omitting behavior in scoring rules, and making the rules and optimal strategies as clear as possible to examinees, is probably preferable when test scores are used to make sensitive placement or selection decisions.

7.2.3 Lord's (1983) Model for Omits

Lord's (1974) treatment of omits as fractionally correct neither presented nor exploited the full likelihood induced by the data. Lord's (1983) model for omits maintains the context of guessing-corrected scoring of multiple-choice items with $A = 1/c$ alternatives, but offers additional structure for the response process. The model first assumes that an examinee either feels a preference for one of the alternatives or is totally undecided among them. The proportion of examinees with ability θ feeling no preference on Item j is $R_j(\theta)$. If a preference is felt, a response is made; and of the responses made by examinees with ability θ who feel a preference, the proportion correct is $P_j^*(\theta)$. If no preference is felt, the examinee will either omit the item with probability ω (an examinee missingness parameter), or respond at random. Note that (θ, ω) constitutes the missingness parameter ϕ in Rubin's notation. Responses and omitting decisions are assumed independent over items, given θ and ω.

This model does not address the correctness or incorrectness of hypothetical responses to omitted items; that is, U_j is undefined when $M_j = 0$. In order to analyze the approach from Rubin's (1976) perspective, we extend it in concert with Lord's hypothesis that an examinee who omits is totally undecided among the alternatives, by positing the following condition:

(vii)   For all items j, $\text{Prob}\left(U_j = 1 \middle| M_j = 0, \theta, \omega\right) = c$ and

$\text{Prob}\left(U_j = 0 \middle| M_j = 0, \theta, \omega\right) = 1 - c$ for all θ and ω;

U_{mis} is thus independent of θ and ω. Assuming Condition (vii), the tree in Figure 3 shows the conditional probabilities of arriving at $\left(m_j, u_j\right)$ in the different possible ways.

[Figure 3 about here]

The likelihood function for the observed data can then be written as follows:

$$\text{Prob}\left(V = \tilde{v} \middle| \theta, \omega\right)$$

$$= \text{Prob}\left(M = \tilde{m} \middle| \theta, \omega\right) \int \prod_{j=1}^{n} \text{Prob}\left(U_j = u_j \middle| \theta, \omega, \tilde{m}_j\right) \partial u_{\text{mis}}$$

$$= \left\{\text{Prob}\left(M = \tilde{m} \middle| \theta, \omega\right)\right\} \left\{\prod_{j:\tilde{m}_j=1} \text{Prob}\left(U_j = \tilde{u}_j \middle| \theta, \omega, M_j = 1\right)\right\} \left\{\int \prod_{j:\tilde{m}_j=0} \text{Prob}\left(U_j = u_j \middle| \theta, \omega, M_j = 0\right) \partial u_{\text{mis}}\right\}$$

$$= \left\{\text{Prob}\left(M = \tilde{m} \middle| \theta, \omega\right)\right\} \left\{\prod_{j:\tilde{m}_j=1} \text{Prob}\left(U_j = \tilde{u}_j \middle| \theta, \omega, M_j = 1\right)\right\} \left\{\int \prod_{j:\tilde{m}_j=0} c_j{}^{u_j} \left(1 - c_j\right)^{1-u_j} \partial u_{\text{mis}}\right\}$$

$$\propto \left\{\text{Prob}\left(M = \tilde{m} \middle| \theta, \omega\right)\right\} \left\{\prod_{j:\tilde{m}_j=1} \text{Prob}\left(U_j = \tilde{u}_j \middle| \theta, \omega, M_j = 1\right)\right\}$$

$$= \left\{\prod_{j=1}^{n} \left[\omega R_j(\theta)\right]^{1-\tilde{m}_j} \left[1 - \omega R_j(\theta)\right]^{\tilde{m}_j}\right\} \left\{\prod_{j:\tilde{m}_j=1} \left[P_j^{**}(\theta, \omega)\right]^{\tilde{u}_j} \left[1 - P_j^{**}(\theta, \omega)\right]^{1-\tilde{u}_j}\right\}, \qquad (7.6)$$

where $P_j^{**}(\theta, \omega) = \text{Prob}\left(U_j = 1 \middle| \theta, \omega, M_j = 1\right)$, or the conditional probability that a response will be correct given that it is observed, is the normalized sum of making a correct response when a preference is felt and of guessing correctly when a preference is not felt but a random response is made:

$$P_j^{**}(\theta, \omega) = \left[1 - \omega R_j(\theta)\right]^{-1} \left\{\left[1 - R_j(\theta)\right]P_j^*(\theta) + c(1 - \omega)R_j(\theta)\right\}.$$

Once V is observed, (7.6) provides a foundation for likelihood and Bayesian inference about θ and ω. Lord suggested that the model could be implemented by specifying functional forms for P_j^* and R_j, such as the 3-parameter logistic IRT function for P_j^* and the 2-parameter logistic with a negative slope for R_j.

THEOREM 7.2.3. Assuming Condition (vii), the model for U_j implicit in Lord's (1983) model for omits can be written in terms of the unidimensional ability variable θ as follows:

$$f_\theta(u_j) = \begin{cases} cR_j(\theta) + [1 - R_j(\theta)]P_j^*(\theta) & \text{if } u_j = 1 \\ (1-c)R_j(\theta) + [1 - R_j(\theta)][1 - P_j^*(\theta)] & \text{if } u_j = 0 \end{cases}. \tag{7.7}$$

Proof.
$$P(U_j = 1|\theta,\omega) = \sum_{k=0}^{1} P(U_j = 1|\theta,\omega,M_j = k)P(M_j = k|\theta,\omega)$$

$$= P(U_j = 1|\theta,\omega,M_j = 1)P(M_j = 1|\theta,\omega) +$$

$$P(U_j = 1|\theta,\omega,M_j = 0)P(M_j = 0|\theta,\omega)$$

$$= \left\{ c(1-\omega)R_j(\theta) + [1 - R_j(\theta)]P_j^*(\theta) \right\} + c\omega R_j(\theta)$$

$$= cR_j(\theta) + [1 - R_j(\theta)]P_j^*(\theta).$$

Note that ω drops out of the expression. Then

$$P(U_j = 0|\theta,\omega) = 1 - P(U_j = 1|\theta,\omega)$$

$$= 1 - cR_j(\theta) - [1 - R_j(\theta)]P_j^*(\theta)$$

$$= (1-c)R_j(\theta) + [1 - R_j(\theta)][1 - P_j^*(\theta)]. \qquad \square$$

Remark. Lord points out that if $\omega=0$ for all examinees (there is no possibility of omitting), then the resulting IRT model is just (7.7). In a manner described by Samejima (1979), the operating characteristic curve for a correct response, or $\text{Prob}(U_j = 1|\theta)$, need not be monotone increasing in θ. Very low ability examinees would feel no preference at all, and answer correctly at a rate equal to c; moderate-ability examinees might tend to feel a preference for a clever distractor and answer correctly at a rate *lower* than c; and high-ability examinees would tend to feel preferences and respond correctly.

Lord's focus on the nonignorable nature of intentional omissions is apparent in the following result.

THEOREM 7.2.4. Consider the case of one item, or $n=1$. (Subscripts on R, P^*, M, etc., may thus be supressed.) Under Lord's (1983) model, augmented by Condition (vii), omitting is ignorable with respect to direct likelihood inference about θ only under the degenerate conditions that either

(a) $\omega = 0$ (i.e., there is no possibility of omitting);

(b)   both $R(\theta)$ and $P^*(\theta)$ are constant with respect to θ (i.e., neither the tendency to feel a preference leading to a response, nor the chances of an observed response being correct, depend on θ); or

(c) $R(\theta) = 0$ (i.e., nobody ever feels a preference among responses).

Proof. If $\omega = 0$, there is no possibility of omitting, so $M \equiv 1$, $U \equiv U_{obs}$, and, for all u_{obs}, $\mathrm{Prob}\big(M = 1 | \tilde{u}_{obs}, \theta, \omega\big) = 1$. Then

$$\mathrm{Prob}\big(V = \tilde{v} | \theta, \omega\big) = \mathrm{Prob}\big(M = 1, U = u_{obs} | \theta, \omega\big)$$
$$= \mathrm{Prob}\big(M = 1 | U = u_{obs}, \theta, \omega\big) \mathrm{Prob}\big(U = u_{obs} | \theta, \omega\big)$$
$$= 1 \times \mathrm{Prob}\big(U = u_{obs} | \theta, \omega\big)$$
$$= \mathrm{Prob}\big(U = u_{obs} | \theta\big),$$

and omitting is (trivially) ignorable.

Now suppose $\omega \neq 0$. Theorem 2.2, NS conditions for ignorability under likelihood inference, requires that the parameter spaces of θ and ϕ are distinct; and that for each ϕ, the expression in (2.8), or $E_{u_{mis}}\big\{g_\phi\big(\tilde{m} | u_{mis}, \tilde{u}_{obs}\big) | \tilde{m}, \tilde{u}_{obs}, \theta, \phi\big\}$, is constant with respect to θ. As for the requirement of distinctness: In general, θ is included in ϕ in Lord's model (that is, $\phi = (\theta, \omega)$), since the probability that an examinee with ability θ will omit Item j is $\omega R_j(\theta)$. If $\omega \neq 0$, then only when $R_j(\theta)$ is constant with respect to θ for all items do the parameter spaces of θ and ϕ become distinct, since ϕ then simplifies to ω alone. Thus, for distinctness to be satisfied for direct-likelihood ignorability when $n = 1$ and $\omega \neq 0$, it is necessary that $R(\theta)$ is constant with respect to θ.

As for the further requirement that (2.8) be constant with respect to θ, we consider separately $\tilde{v} = (0, *)$, $(1,1)$, and $(1,0)$, since for ignorability to hold in general, it must hold for each potential pattern of observations. The required expressions are derived from Figure 3.

$\underline{(0,*)}$: If $\tilde{v} = (0, *)$, then

$E_{u_{mis}}\big\{g_\phi\big(\tilde{m} | u_{mis}, \tilde{u}_{obs}\big) | \tilde{m}, \tilde{u}_{obs}, \theta, \phi\big\}$
$\quad = \mathrm{Prob}\big(M = 0 | U = 0, \theta, \omega\big) \mathrm{Prob}\big(U = 0 | \theta, \omega\big) + \mathrm{Prob}\big(M = 0 | U = 1, \theta, \omega\big) \mathrm{Prob}\big(U = 1 | \theta, \omega\big)$
$\quad = \mathrm{Prob}\big(M = 0, U = 0 | \theta, \omega\big) + \mathrm{Prob}\big(M = 0, U = 1 | \theta, \omega\big)$
$\quad = R(\theta)\omega(1 - c) + R(\theta)\omega c$
$\quad = R(\theta)\omega.$

This expression is in fact constant in θ if $R(\theta)$ is constant in θ, so ignorability holds under this condition if the observation is an omitted response. We now see, however, that this is not sufficient when the observation is an observed right or wrong response.

(1,1) and (1,0): If $\tilde{v} = (1,1)$, then there is no u_{mis} to integrate over and

$$E_{u_{mis}}\left\{g_\phi\left(\tilde{m}|u_{mis},\tilde{u}_{obs}\right)|\tilde{m},\tilde{u}_{obs},\theta,\phi\right\} = \text{Prob}\left(M = 1|U = 1,\theta,\omega\right)$$

$$= \frac{\text{Prob}\left(M = 1,U = 1|\theta,\omega\right)}{\text{Prob}\left(U = 1|\theta,\omega\right)}$$

$$= \frac{[1 - R(\theta)]P^*(\theta) + cR(\theta)(1 - \omega)}{[1 - R(\theta)]P^*(\theta) + cR(\theta)}. \qquad (7.8a)$$

Similarly, if $\tilde{v} = (1,0)$,

$$E_{u_{mis}}\left\{g_\phi\left(\tilde{m}|u_{mis},\tilde{u}_{obs}\right)|\tilde{m},\tilde{u}_{obs},\theta,\phi\right\} = \text{Prob}\left(M = 1|U = 0,\theta,\omega\right)$$

$$= \frac{\text{Prob}\left(M = 1,U = 0|\theta,\omega\right)}{\text{Prob}\left(U = 0|\theta,\omega\right)} \qquad (7.8b)$$

$$= \frac{[1 - R(\theta)]\left[1 - P^*(\theta)\right] + (1 - c)R(\theta)(1 - \omega)}{[1 - R(\theta)]\left[1 - P^*(\theta)\right] + (1 - c)R(\theta)}.$$

Given that $\omega \neq 0$, both (7.8a) and (7.8b) are constant with respect to θ if either $R(\theta)=0$ or both $R(\theta)$ and $P^*(\theta)$ are constant with respect to θ (the degenerate case in which the item conveys no information whatsoever about θ). $\qquad\square$

7.2.4 Nominal Category Models

IRT models for multiple-category items give probabilities of item responses as conditionally independent functions of θ (e.g., Bock, 1972; Samejima, 1979; Thissen & Steinberg, 1986). These models have sometimes been used for data with intentional omissions, with an omit treated as one more possible response to a multiple-choice item. Lord (1983) expresses reservations about this practice, "... since it treats probability of omitting as dependent only on the examinee's ability, whereas it actually depends on a dimension of temperament. It seems likely that local unidimensional independence may not hold" (p. 477). The following analysis makes Lord's concerns more explicit.

The character of this approach regarding omission are seen most easily in the case in which all incorrect overt responses are denoted into a single category. Suppose the following conditions hold:

(viii)   A SMURFLI[2] IRT model in θ governs U;

(ix) ϕ characterizes an examinee's tendency to omit, via $g_\phi(m_j | u_j)$, and

(x) (u_j, m_j) are conditionally independent over items given (θ, ϕ); that is,

$$\text{Prob}(U = u, M = m | \theta, \phi) = \prod_j \text{Prob}(U_j = u_j, M_j = m_j | \theta, \phi),$$

which in turn implies

$$\text{Prob}(V = v | \theta, \phi) = \prod_j \text{Prob}(V_j = v_j | \theta, \phi). \tag{7.9}$$

Recalling that the values 0, 1, and * of v indicate observed wrong, observed right, and omit, we obtain the multiple-category model probabilities f_θ^* as follows:

$$f_\theta^*(V_j = 0) = \text{Prob}(U_j = 0, M_j = 1 | \theta)$$
$$= \int f_\theta(U_j = 0) g_\phi(M_j = 1 | U_j = 0) p(\phi | \theta) \partial\phi, \tag{7.10a}$$

$$f_\theta^*(V_j = 1) = \text{Prob}(U_j = 1, M_j = 1 | \theta)$$
$$= \int f_\theta(U_j = 1) g_\phi(M_j = 1 | U_j = 1) p(\phi | \theta) \partial\phi, \tag{7.10b}$$

and

$$f_\theta^*(V_j = *) = \text{Prob}(U_j = 0, M_j = 0 | \theta) + \text{Prob}(U_j = 1, M_j = 0 | \theta)$$
$$= \int f_\theta(U_j = 0) g_\phi(M_j = 0 | U_j = 0) p(\phi | \theta) \partial\phi \tag{7.10c}$$
$$+ \int f_\theta(U_j = 1) g_\phi(M_j = 0 | U_j = 1) p(\phi | \theta) \partial\phi.$$

THEOREM 7.2.5. Under Conditions (viii)-(x) above, the multiple-category response functions in θ for v, or (7.10a)-(7.10c), exhibit conditional independence over items only if all examinees with each given θ have the same value of ϕ.

Proof. The conditional probability of V_j given θ, marginal with respect to ϕ, is

$$\text{Prob}(V_j = v_j | \theta) = \int \text{Prob}(V_j = v_j | \theta, \phi) p(\phi | \theta) \partial\phi.$$

If local independence given θ is to hold *across* items, it follows that

$$\text{Prob}(V = v | \theta) = \prod_j \int \text{Prob}(V_j = v_j | \theta, \phi) p(\phi | \theta) \partial\phi. \tag{7.11}$$

But marginalizing over ϕ in $\text{Prob}(V = v | \theta, \omega)$ to obtain $\text{Prob}(V = v | \theta)$ yields

$$\text{Prob}(V = v | \theta) = \int \text{Prob}(V = v | \theta, \phi) p(\phi | \theta) \partial\phi$$
$$= \int \prod_j \text{Prob}(V_j = v_j | \theta, \phi) p(\phi | \theta) \partial\phi. \tag{7.12}$$

Expressions (7.11) and (7.12) differ only in that the order of integration and multiplication are interchanged, and are equal in general only if ϕ is constant for each given value of θ. $\square$

8.0   Examinee Choice of Items

The instructions to one section of the College Entrance Examination Board's 1905 German test read, " Answer only *one* of the following six questions," and the 1905 Botony exam included a section of ten items, of which students had to answer any seven (Wainer & Thissen, 1994, p. 161). In this section we consider tests that (a) incorporate choice in such a manner and (b) posit a SMURFLI[2] IRT model. These tests present problems of inference with respect to the missing responses to non-chosen items that are formally identical to those associated with intentional omissions: The students examines all items, consider their chances of success on each, and chose which to answer in a manner that may depend on their actual chances of success through θ. The only difference is the constraint on possible missingness patterns, which is irrelevant to Rubin's ignorability conditions. Section 8.1 recasts the results of Section 7.1 in the context of examinee choice of items, concluding that if an underlying SMURFLI[2] IRT model is to be maintained, it is necessary to model missingness and response jointly. Section 8.2 discusses how the generally-nonignorable missingness can be modeled in the IRT context.

8.1   On the Ignorability of Responses to Non-Chosen Items

As with intentional omits, motivated examinees facing choice situations attempt to make choices and responses that maximize the score they expect to receive. We assume a SMURFLI[2] IRT model, and consider test scores in the form of counts of correct response $R(v)$. Since a correct response to an item gives a higher test score than an incorrect response, examinees who want to obtain high scores will chose an allowable missingness pattern in which they can make responses they think are likely to be correct. As with omits, examinees' perceived probabilities of correct response are not necessarily the same as those given by the IRT model, and examinees may differ in the accuracy of their estimates. Such characteristics of an examinee constitute the choice parameter ϕ. The intuition, supported by studies such as that of Wang, Wainer, and Thissen (1993), again suggests that examinees with high abilities are typically better at projecting their expected scores under different choice patterns and responding to item configurations that lead them to higher scores.

THEOREM 8.1.1. For a given $\tilde{m}$ and $\tilde{u}_{\text{obs}}$, the missingness process induced by examinee choice is MAR only if this missingness pattern is equally likely for all values of u_{mis}.

Proof. This follows immediately from the definition of MAR. $\qquad\qquad\square$

Remark. If examinees are more likely to avoid choosing items when they think their responses would lead to lower scores, MAR implies the untenable belief of independence of their perceptions of correctness from the probabilities given by the IRT model.

Assuming a monotone unidimensional IRT model and considering a given missingness pattern $\tilde{m}$ induced by examinee choice, the NS conditions for ignorability in Theorems 2.2 and 2.3 say that for the probability of $\tilde{m}$ is the same for all θ even though the expected number-right score of the non-chosen items, or $E\{R(U_{\text{mis}})|\theta\}$, is increasing in θ (by Theorem 7.1.1). That is, high-ability examinees are just as likely to produce this choice pattern as low-ability examinees (given $\tilde{u}_{\text{obs}}$), even though the associated responses have a greater expected contribution to their total test score and the maximum 'chosen M' would be higher. A counter-intuitive result is easiest to see when $n=2$:

THEOREM 8.1.2. Suppose a SMURFLI[2] IRT model holds in a domain consisting of two items ($j=1,2$), and the missingness induced by "answer any one of two" format is ignorable under likelihood or Bayesian inference. Suppose further that $U_1=0$. Then (a) the conditional probability that an examinee will choose to present this response rather than the response to U_2 must be constant with respect to θ, even though (b) Prob($U_2=1$) is increasing with respect to θ.

Proof. The NS condition for ignorability under likelihood inference (2.8) requires that for each value of ϕ, the value of $E_{u_{\text{mis}}}\left\{g_\phi\left(\tilde{m}|u_{\text{mis}},\tilde{u}_{\text{obs}}\right)|\tilde{m},\tilde{u}_{\text{obs}},\theta,\phi\right\}$ (the probability that missingness pattern $\tilde{m}$ will be observed given $\tilde{u}_{\text{obs}}$, ϕ, and θ) is constant with respect to θ. The theorem addresses the case in which $n=2$, $\tilde{m}=(1,0)$, and $\tilde{v}=(0,*)$. Because the IRT model is strictly monotonic, the probability of the missing response being correct is increasing in θ. The same argument holds for ignorability under Bayesian inference, since, by Theorem 2.3, it is required that the expectation of (2.8) over ϕ, or (2.9), is constant with respect to θ. $\qquad\qquad\square$

8.2 Modeling Examinee Choice Within the IRT Framework

This section addresses the information in examinee-choice test responses about θ. Because missingness due to examinee choice is not generally ignorable, appropriate inference under the IRT framework requires working from v rather than simply u_{obs}. Wainer and Thissen (1994) point out that as with omits, the correct model cannot be ascertained from observations of v alone. Supplemental data collections of (u,m), as was done in Wang, Wainer, and Thissen (1993) suffice to build a model. This study asked examinees to respond to both of two items and to also indicate which one they would have chosen if only one were to have been scored. From such data it is possible to estimate both $f_\theta(u)$, the IRT model, and $g_\phi(m|u)$, the choice model. Note that g must assign a probability of zero to any m other than the $(n$-choose-$N)=n!/[N!(n-N)!]$ patterns consistent with "answer any N of n" format.

THEOREM 8.2.1. The likelihood function for induced θ under an LI IRT model when responses are observed under the "answer any N of n" format is the weighted average of all possible response vectors u consistent with $\tilde{u}_{obs}$, with the weights associated with each pattern of non-observed responses being the probabilities of the observed choice pattern given u_{obs} and the non-observed responses.

Proof. This is a specialization of (2.2) to the context of choice. The likelihood function induced by $\tilde{v}$ is $L(\theta,\phi|\tilde{v}) = \delta_{\theta\phi} \int f_\theta(u_{mis},\tilde{u}_{obs}) g_\phi(\tilde{m}|u_{mis},\tilde{u}_{obs}) du_{mis}$. $\qquad\square$

8.2.1 Special Cases

This section gives $g_\phi(\tilde{m}|u_{mis},\tilde{u}_{obs})$ for some special cases of choice behavior with SMURFLI[2] items and "answer any N of n" format.

SPECIAL CASE #1: *An MCAR choice mechanism.* If an examinee's choice of items is independent of θ, then g assigns equal probability to all m consistent with the instructions and zero to all others:

$$g_\phi(m|u) = \begin{cases} K^{-1} & \text{if } \sum_j m_j = N \\ 0 & \text{otherwise,} \end{cases} \tag{8.1}$$

where $K=(n$-choose-$N)$. The likelihood under ignorability can be expressed as the equally-weighted average of all complete response patterns consistent with $\tilde{u}_{obs}$:

$$L(\theta|\tilde{v}) = \delta_\theta \int f_\theta(u_{mis},\tilde{u}_{obs}) K^{-1} du_{mis}, \tag{8.2}$$

which is proportional to $L^*(\theta | \tilde{u}_{obs})$. In this situation, $\tilde{u}_{obs}$ is representative of an examinee's *typical* performance, despite the choice format.

SPECIAL CASE #2: *Fully efficient choice, unique maximum score.* Suppose an examinee has perfect knowledge of $\tilde{u}$ and chooses $\tilde{m}$ in such a way as to maximize the test score, $T(v)$. If the highest possible score is obtained by a unique pattern $v^* = \left(u^*_{obs}, m^*\right)$, then

$$g_\phi\left(m | u_{mis}, u_{obs}\right) = \begin{cases} 1 & \text{if } T(v) = T\left(v^*\right) \text{ and } \sum_j m_j = N \\ 0 & \text{otherwise.} \end{cases} \tag{8.3}$$

By Theorem 8.2.1, $L(\theta, \phi | \tilde{v})$ in this case is an equally weighted average of all $f_\theta\left(u_{mis}, \tilde{u}_{obs}\right)$ possibilities in which any other choice pattern would have yielded a lower score than $T\left(v^*\right)$. In this case and the next, $\tilde{u}_{obs}$ represents *maximal* performance.

SPECIAL CASE #3: *Fully efficient choice, nonunique maximum score.* Suppose that an examinee has perfect knowledge of $\tilde{u}$ and chooses $\tilde{m}$ in such a way as to maximize the test score, $T(v)$. If $v^* = \left(u^*_{obs}, m^*\right)$ is one of H patterns that yields the highest score from $\tilde{u}$ and the examinee chooses one of these patterns at random, then

$$g_\phi\left(m | u_{mis}, u_{obs}\right) = \begin{cases} H^{-1} & \text{if } T(v) = T\left(v^*\right) \text{ and } \sum_j m_j = N \\ 0 & \text{otherwise.} \end{cases} \tag{8.4}$$

SPECIAL CASE #4: *Partially efficient choice.* Suppose that an examinee seeks to choose $\tilde{m}$ in such a way as to maximize the test score, $T(v)$, but has imperfect knowledge of $\tilde{u}$. One specification of g that takes this intention into account while softening its effect is a linear combination of the ignorability weights (8.1) and the efficient-choice weights, namely (8.3) or (8.4) as appropriate. For a non-unique maximum score, for example,

$$g_\phi\left(m | u_{mis}, u_{obs}\right) = \begin{cases} \phi H^{-1} + (1-\phi)K^{-1} & \text{if } T(v) = T\left(v^*\right) \text{ and } \sum_j m_j = N \\ (1-\phi)K^{-1} & \text{if } T(v) < T\left(v^*\right) \text{ and } \sum_j m_j = N \\ 0 & \text{otherwise,} \end{cases} \tag{8.5}$$

with $0 < \phi < 1$. There are K patterns with N observed responses, H of which yield the maximal score and K-H of which yield lower scores; it can be verified that the probabilities in (8.5) sum to 1 over the K allowable patterns. Efficient-choice weights that presume $\tilde{u}_{obs}$

represents "maximum performance" are thus mixed with ignorability weights that presume it represents "typical performance".

Remark. Special Case #4 might apply to a not-uncommon practice with IRT: the administration of "customized tests," in which a school or district selects items from an item bank for which item parameters have been estimated from a non-targeted reference population of examinees (Yen, Green, & Burket, 1987). The items of a customized test are chosen because they are more relevant to the local curriculum than the non-chosen items, so the administrator is effectively acting as an agent of the local students in a choice exam. If this choice is ignored, over-estimates of θ result in a manner shown in following example.

8.2.2 A Numerical Example

This example concerns "answer any N of n" format with iid SMURFLI[2] items following an IRT model of the form $f_\theta(u_j) = .2 + .8u_j \exp(\theta)/[1 + \exp(\theta)]$, with $p(\theta) = N(0,2)$ and number-right scoring. In particular, we consider $n=5$ and $N=2$, for $\tilde{v} = (1,1,*,*,*)$ and $\tilde{v} = (1,0,*,*,*)$ under three alternatives to modeling the choice process.

The first panel in Figure 4 gives the likelihood functions for the eight possible Us consistent with $\tilde{v} = (1,1,*,*,*)$. $L(\theta|\tilde{v})$ is a weighted average of these, with weights depending on how the choice process is modeled. Table 1 gives weights $g_\phi(m|u)$ under ignorability, efficient choice (8.4), and a 9:1 mixture of efficient-choice and ignorability weights via (8.5). The ignorability weights are all (5-choose-2)$^{-1}$, or .10. For efficient choice, $g_\phi(m|u)$ depends on the number of different u_{obs} with a score of 2 that could be made from the complete pattern, or $H(u)$. When $u = (1,1,0,1,1)$, for example, $H(u)=$ (4-choose-2)=6, so $H(u)^{-1}=.167$ appears in the efficient-choice column for this pattern. The resulting $L(\theta|\tilde{v})$s appear in the second panel of Figure 4. The likelihood under efficient choice flattens out above 0, whereas the ignorability likelihood continues to rise rapidly; examinees above zero are likely to obtain at least two 1's in U, so this observational scheme provides little evidence for distinguishing among them. The distributions $p(\theta|\tilde{v})$ appear in the final panel of Figure 4, and the posterior mean and standard deviation for θ in each case appear in Table 1. The flattened likelihood for the efficient-choice condition leads to a much lower posterior mean and a somewhat wider posterior standard deviation than are obtained under ignorability.

[Table 1 & Figure 4 about here]

Similar results for $\tilde{v} = (1,0,*,*,*)$ appear in Table 2 and Figure 5. Efficient-choice again yields a much lower posterior mean than the ignorability condition, but now, interestingly, it leads to a smaller posterior standard deviation. This is because under efficient choice, observing u_{obs} with only one correct response means that the three missing responses must all be zero; i.e., under efficient-choice, observing $\tilde{v} = (1,0,*,*,*)$ is equivalent to observing $u = (1,0,0,0,0)$.

[Table 2 & Figure 5 about here]

In this example, examinee-choice makes inference about θ less efficient for higher values, where the question is about the *average value* of u_js, but more efficient for very low values where the question is more one of the *existence* of any u_js that are 1. Departure from efficient-choice behavior erodes this potential advantage. Results such as these led both Wainer and Thissen (1994) and Bradlow and Thomas (in review) to conclude that examinee-choice and IRT-based inference are an unattractive combination. IRT-based inference works well when one is interested in typical performance in a domain of exchangeable tasks, to be characterized by a single ability variable θ; examinee-choice is attractive when it is not typical performance in a domain of exchangeable tasks, but the maximal performance in special circumstances best known to the examinee. An alternative modeling approach for such circumstances requires the specification of a common framework of evaluation into which performances in different contexts can be mapped (e.g., the Advanced Placement Studio Art portfolio assessment, in which students have considerable choice as to media, style, and subjects; see Myford & Mislevy, 1995).

9.0 Summary

In practical applications of item response theory (IRT), there are several reasons that item responses may not be observed from all examinees to all test items. We used Rubin's (1976) theorems to determine whether ignorability holds under direct likelihood and Bayesian inference about examinee parameters θ under six common types of missingness in IRT, with item parameters known. Ignoring the missingness process under direct likelihood inference means using a pseudo-likelihood that includes terms for only the responses that were observed, without regard for the processes by which they came to be observed. The resulting inferences are appropriate if the pseudo-likelihood is proportional to the correct likelihood that does account for the missingness process. The missingness is ignorable with respect to Bayesian inference if the correct posterior is proportional to the

product of the pseudo-likelihood and an appropriate prior distribution. Our findings are summarized below. Table 3 highlights the results on ignorability.

[Insert Table 3 about here]

Alternate test forms. When an examinee is assigned one of several alternative test forms by a random process such as a coin flip or a spiraling scheme, the process that renders missing the responses to items on the forms not presented is ignorable under both likelihood and Bayesian inference for θ.

Targeted testing. When covariates such as educational or demographic status are used to assign an examinee one of several tests that differ in their measurement properties, the resulting missingness on forms not given is ignorable under direct likelihood inference for θ, but not under Bayesian inference unless the prior information about examinees that led to differential assignments is conditioned on.

Adaptive testing. The missingness caused by the selection of items to present to an examinee based on observed responses to previous items is ignorable under both direct likelihood and Bayesian inference. It should be noted that ignorability under direct likelihood inference means that the correct points are identified as MLEs of θ, but the usual MLE properties under sampling-distribution inference need not hold because the probabilities of missingness patterns depend on the values of observed responses.

Not-reached items. When some examinees do not interact with the last items on a nearly unspeeded test, the not-reached process is ignorable with respect to direct likelihood inference about θ. This missingness process is not ignorable under Bayesian inference unless speed and ability are independent.

Omitted items. When examinees are presented items, appraise their content, and decide for their own reasons not to respond, the missingness is not generally ignorable. Inferences must be drawn from a full model for the joint distribution of missingness and item response, as sketched in Lord (1983). Under the assumption that examinees are perfect judges of their chances of responding correctly, and omit only if it is in accordance with the strategy that maximizes their expected score, Lord's (1974) treatment of omits as fractionally correct under a standard IRT model can be justified as providing the expectation of a conditional term in the full likelihood. This procedure is readily incorporated into standard complete-data IRT algorithms and avoids having to specify the full likelihood, but

foregoes information about examinee and item parameters conveyed by the observed pattern of missingness.

Examinee choice of items. Insofar as ignorability conditions are concerned, examinee choice is equivalent to intentional omission. Choice is not generally ignorable, and treating it as such typically overestimates the information about θ. This includes choice made by a test administrator on behalf of examinees—e.g., school officials pick out a "customized test" aligned to their curriculum, but using item parameters estimated from a non-targeted reference population. It is possible, with supplemental data from experiments, to estimate the typically lower and more diffuse likelihoods for θ arising from choice in IRT domains, but the administration scheme of examinee choice of tasks is ill-suited to domain-referenced IRT inference. Conditional evaluation of choice performances within a common framework of evaluation seems better suited to tasks that evidence targeted skills only given ancillary skills.

References

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Bradlow, E.T., & Thomas, N. (in review). Item response theory models applied to data allowing examinee choice. Princeton, NJ: Educational Testing Service.

Chang, H., & Ying, Z. (in preparation). Recursive maximum likelihood estimation, sequential design, and computerized adaptive testing. Princeton, NJ: Educational Testing Service.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Holland, P.W., & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent trait variable models. *Annals of Statistics, 14*, 1523-1543.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lord, F.M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39*, 247-264.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Lord, F.M. (1983). Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika, 48*, 477-482.

Mislevy, R.J., & Wu, P-K. (1988). *Inferring examinee ability when some item responses are missing.* Research Report RR-88-48-ONR. Princeton, NJ: Educational Testing Service.

Myford, C.M., & Mislevy, R.J. (1995). *Monitoring and improving a portfolio assessment system.* Center for Performance Assessment Research Report. Princeton, NJ: Educational Testing Service.

Owen, R.J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association, 70*, 351-356.

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Sabers, D.L., & Feldt, L.S. (1968). An empirical study of the effect of the correction for chance success on the reliability and validity of an aptitude test. *Journal of Educational Measurement, 5*, 251-258.

Samejima, F. (1979). A new family of models for the multiple-choice item. *Research Report 79-4*. Knoxville, TN: Department of Psychology, University of Tennessee.

Sherriffs, A.C., & Boomer, D.S. (1954). Who is penalized by the penalty for guessing? *Journal of Educational Psychology, 45*, 81-90.

Snijders, J.T., Souren, C.J.M.H., & Welten, V.J. (1963). *ISI- publikaties 1. Algemene Handleiding*. Groningen: Wolters.

Stocking, M.L., Eignor, D., & Cook, L. (1988). Factors affecting the sample invariant properties of linear and curvilinear observed and true score equating procedures. *Research Report RR-88-41*. Princeton, NJ: Educational Testing Service.

Thissen, D., & Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika, 49*, 501-519.

van den Wollenberg, A.L. (1979). *The Rasch model and time-limit tests*. Doctoral dissertation, University of Nijmegen.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wainer, H., & Thissen, D. (1994). On examinee choice in educational testing. *Review of Educational Research, 64*, 159-195.

Wang, X.B., Wainer, H., & Thissen, D. (1995). On the viability of some untestable assumptions in equating exams that allow examinee choice. *Applied Measurement in Education, 8*, 211-225.

Wolf, R. (1977). *Achievement in America*. New York: Teachers College Press.

Yamamoto, K., & Everson, H.T. (1995). Modeling the mixture of IRT and pattern responses by a modified Hybrid model. *ETS Research Report RR-95-16*. Princeton, NJ: Educational Testing Service.

Yen, W.M., Green, D.R., & Burket, G.R. (1987). Valid normative information from customized achievement tests. *Educational Measurement: Issues and Practice, 6*, 7-13.

Zwick, R. (1990). When do item response function and Mantel-Haenzel definitions of differential item functioning coincide? *Journal of Educational Statistics, 15*, 185-197.

Table 1
Weighting Factors for Likelihoods when $\tilde{v} = (1,1,*,*,*)$

$u\left[=\left(1,1,u_{\text{mis}}\right)\right]$	$g_\phi(m\|u)$		
	Ignorability	Fully-efficient Choice	Partially-Efficient Choice, $\phi = .9$
11000	.10	1.00	.91
11001	.10	.33	.31
11010	.10	.33	.31
11011	.10	.17	.16
11100	.10	.33	.31
11101	.10	.17	.16
11110	.10	.17	.16
11111	.10	.10	.10
$E(\theta\|\tilde{v})$	1.32	.52	.82
$Var(\theta\|\tilde{v})$	1.56	1.72	1.71

Table 2
Weighting Factors for Likelihoods when $\tilde{v} = (1,0,*,*,*)$

$u\left[=\left(1,0,u_{\text{mis}}\right)\right]$	$g_\phi(m\|u)$		
	Ignorability	Fully-efficient Choice	Partially-Efficient Choice, $\phi = .9$
10000	.10	1	.91
10001	.10	0	.01
10010	.10	0	.01
10011	.10	0	.01
10100	.10	0	.01
10101	.10	0	.01
10110	.10	0	.01
10111	.10	0	.01
$E(\theta\|\tilde{v})$	-.52	-1.84	-1.31
$Var(\theta\|\tilde{v})$	1.50	1.20	1.48

Table 3
Ignorability Results for Estimating θ Given Item Parameters

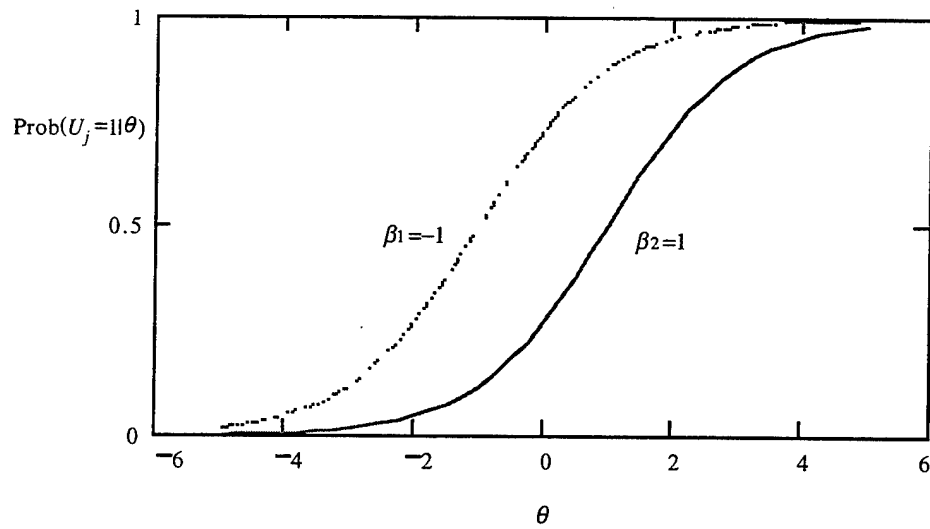| Type of Missingness	Type of Inference	
	Direct Likelihood	Bayesian
Alternate Test Forms	Yes	Yes
Targeted Test Forms	Yes	Yes, conditional on examinee covariates
Adaptive Testing	Yes	Yes
Not-Reached Items	Yes	No, unless speed and ability are independent
Intentional Omits	No	No
Examinee Choice	No	No

Figure 1

Response Curves for Correct Response to Two
Dichotomous Rasch-Model Items
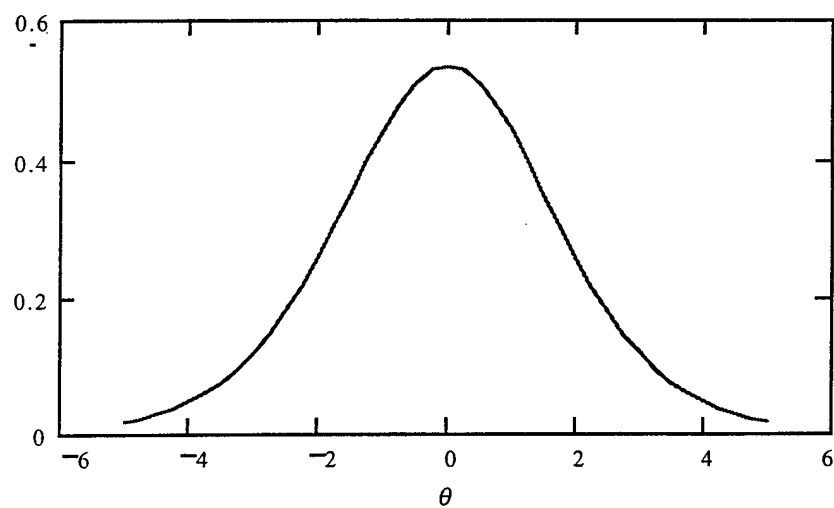
Likelihood
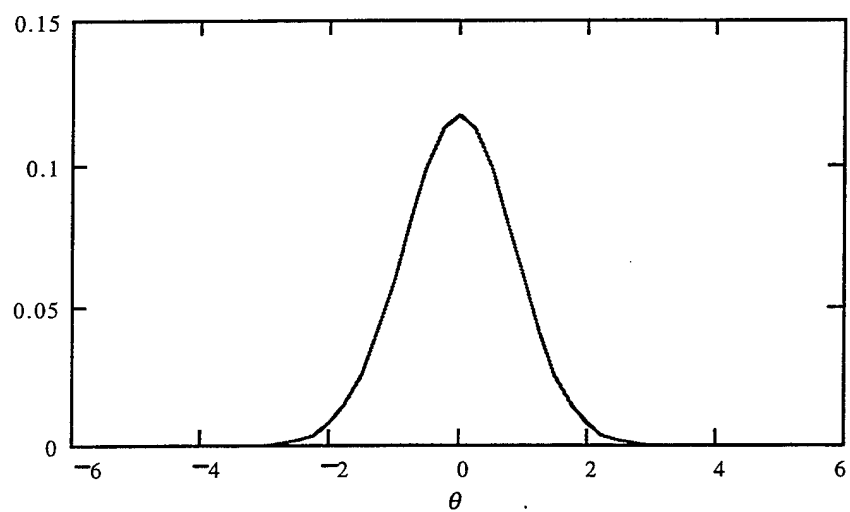
Posterior

Figure 2

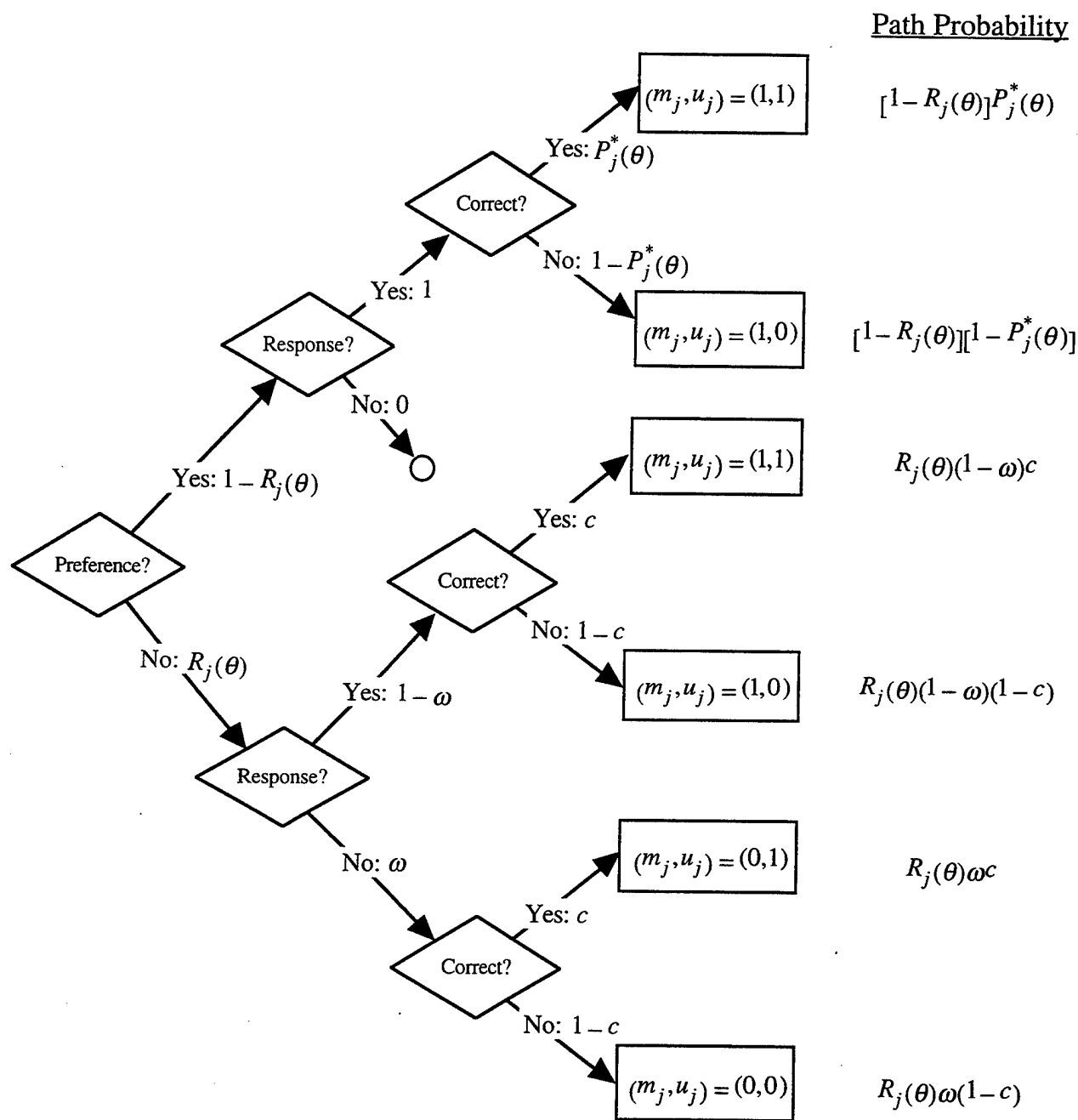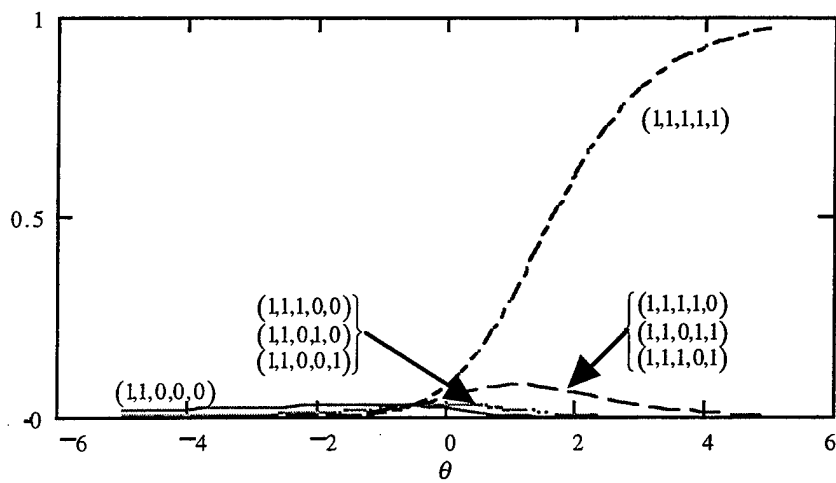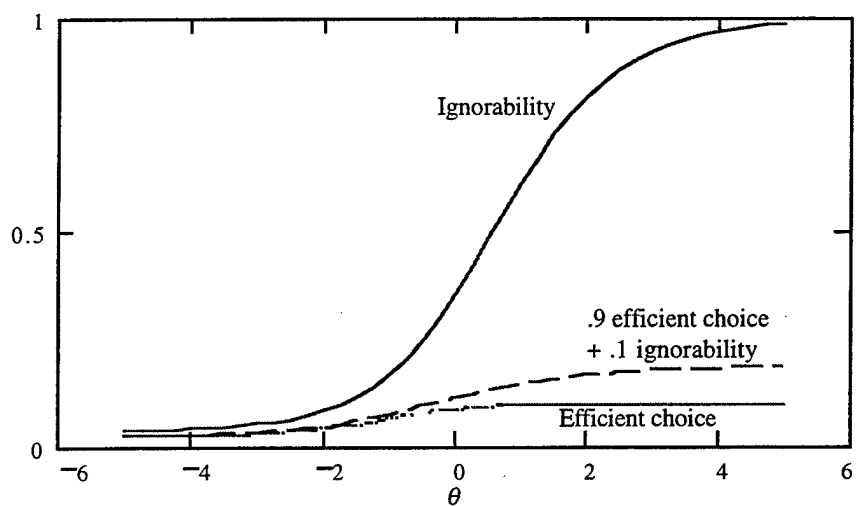Likelihood and Posterior Distribution for θ after Observing $u=(1,0)$

Path Probability

$[1 - R_j(\theta)]P_j^*(\theta)$

$[1 - R_j(\theta)][1 - P_j^*(\theta)]$

$R_j(\theta)(1 - \omega)c$

$R_j(\theta)(1 - \omega)(1 - c)$

$R_j(\theta)\omega c$

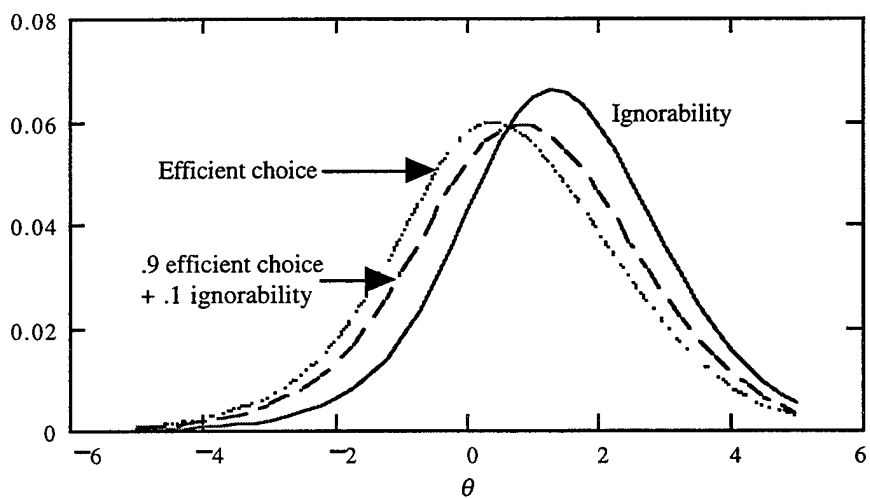$R_j(\theta)\omega(1 - c)$

Figure 3

Conditional Probabilities of Possible Paths to Responses

Likelihood functions for θ given us
that are consistent with u_{obs}

Likelihood functions for θ given v under different
assumptions about the choice process

Posterior distributions for θ given v under different
assumptions about the choice process

Figure 4

Inference about θ from $v=(1,1,*,*,*)$

Likelihood functions for θ given us
that are consistent with u

Likelihood functions for θ given v under different
assumptions about the choice process

Posterior distributions for θ given v under different
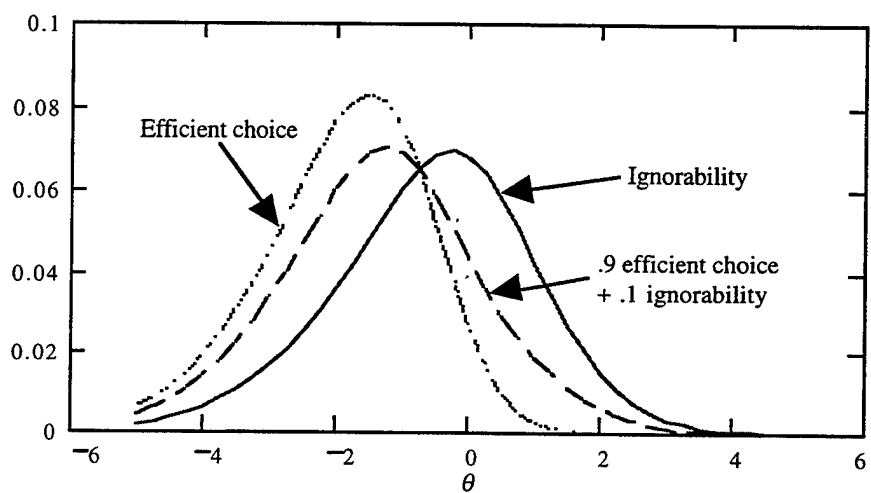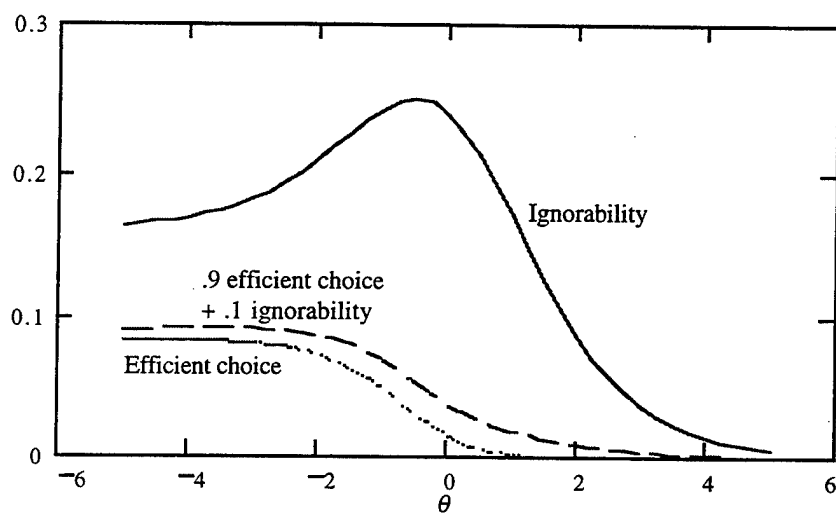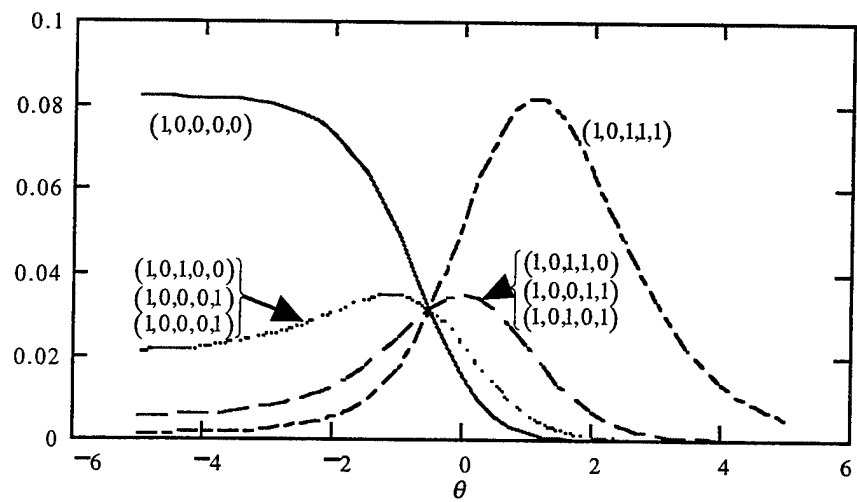assumptions about the choice process

Figure 5

Inference about θ from $v=(1,0,*,*,*)$

Distribution List

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Robert L Albright
Educational Testing Service
16-C Rosedale Road
Princeton, NJ 08541

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
School of Education
Tel Aviv University
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Dr Nancy S Cole
Educational Testing Service
14-C Rosedale Road
Princeton NJ 08541

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
  Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
    Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education Research
Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr. George Macready
Professor
Dept of Meas., Stat. & Eval.
College of Education
Room 1230C, Benjamin Bldg.
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA   22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE   19716

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
   in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC   20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
2201 N. Dodge Street
P. O. Box 168
Iowa City, IA 52243

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA   94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis A Roussos
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Stephen Sireci
American Council on Education
GED Testing Service Suite 250
One Dupont Circle NW
Washington DC 20036

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550